



(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
11.03.2020 Bulletin 2020/11

(51) Int Cl.:
C12N 15/113 (2010.01) **A61K 48/00** (2006.01)
C12N 15/85 (2006.01) **C12N 15/867** (2006.01)

(21) Application number: **19185730.9**

(22) Date of filing: **04.09.2015**

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

(30) Priority: **04.09.2014 US 201462045997 P**

(62) Document number(s) of the earlier application(s) in accordance with Art. 76 EPC:
15837740.8 / 3 189 143

(71) Applicants:
• **Memorial Sloan-Kettering Cancer Center**
New York, NY 10065 (US)
• **University of Washington**
Seattle, WA 98105 (US)

(72) Inventors:
• **SADELAIN, Michel**
New York, NY 10024 (US)
• **RIVIERE, Isabelle**
New York, NY 10024 (US)

• **MANSILLA-SOTO, Jorge**
Forest Hills, NY 11375 (US)
• **WANG, Xiuyan**
New York, NY 10065 (US)
• **STAMATOYANNOPOULOS, George**
Seattle, WA 98105 (US)
• **STAMATOYANNOPOULOS, John**
Seattle, WA 98105 (US)
• **LIU, Mingdong**
Seattle, WA 98105 (US)

(74) Representative: **Barker Brettell LLP**
100 Hagley Road
Edgbaston
Birmingham B16 8QQ (GB)

Remarks:

- This application was filed on 11-07-2019 as a divisional application to the application mentioned under INID code 62.
- Claims filed after the date of receipt of the divisional application (Rule 68(4) EPC).

(54) **GLOBIN GENE THERAPY FOR TREATING HEMOGLOBINOPATHIES**

(57) The presently disclosed subject matter provides for expression cassettes that allow for expression of a globin gene or a functional portion thereof, vectors comprising thereof, and cells transduced with such expres-

sion cassettes and vectors. The presently disclosed subject matter further provides methods for treating a hemoglobinopathy in a subject comprising administering an effective amount of such transduced cells to the subject.

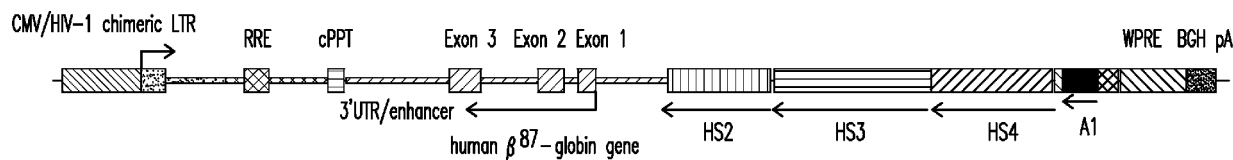


FIG. 1

Description**CROSS-REFERENCE TO RELATED APPLICATIONS**

- 5 **[0001]** This application claims priority to United States Provisional Application No. 62/045,997, filed September 4, 2014, the contents of which is hereby incorporated by reference in its entirety herein.

GRANT INFORMATION

- 10 **[0002]** This invention was made with government support under Grant No. HL053750 from National Heart, Lung and Blood Institute. The government has certain rights in the invention.

INTRODUCTION

- 15 **[0003]** The presently disclosed subject matter provides expression cassettes and vectors comprising such expression cassettes that express a globin protein, e.g., a human β -globin protein. The presently disclosed subject matter further provides expression cassettes that comprise a globin gene or a functional portion thereof operably linked to a β -globin locus control region (LCR) comprising a plurality of Dnase I hypersensitive sites. The expression cassettes of the presently disclosed subject matter comprise one or more insulators that counteract the effect of enhancer elements. The insulators
20 disclosed herein do not substantially adversely impact the titer of a vector that comprises the presently disclosed expression cassettes. The expression cassettes and vectors can be used for treating a hemoglobinopathy, e.g., β -thalassemia, and sickle cell anemia.

BACKGROUND

- 25 **[0004]** β -thalassemia and sickle cell anemia are severe congenital anemias that are caused by defective production of the β chain of hemoglobin. In β -thalassemia, the β chain deficit leads to the intracellular precipitation of excess α -globin chains, causing ineffective erythropoiesis and hemolytic anemia (Weatherall and Clegg (1981), Stamatoyannopoulos *et al.*, (1994), Weatherall (2001), Steinberg (2001)). In the most severe forms found in homozygotes or compound heterozygotes, anemia is lethal within the first years of life in the absence of any treatment (Cooley and Lee (1925)).
30 Lifelong transfusion therapy is needed to correct anemia, suppress ineffective erythropoiesis and inhibit gastrointestinal iron absorption (Weatherall and Clegg (1981), Stamatoyannopoulos *et al.* (1994), Weatherall (2001), Steinberg (2001)). However, transfusion therapy itself leads to iron overload, which is lethal if untreated. The prevention and treatment of iron overload are the major goals of current patient management (Giardina (2001)). The only current curative treatment
35 to cure β -thalassemia is to provide erythroid precursors harboring normal globin genes through allogeneic bone marrow transplantation (BMT) (Giardini and Lucarelli (1994), Boulad *et al.* (1998), Lucarelli *et al.* (1999), Tisdale and Sadelain (2001)).

- [0005]** In sickle cell anemia, the hemoglobin β chain is mutated at amino acid position 6 (Glu $\rightarrow$ Val), leading to the synthesis of β^S instead of the normal β^A chain (Steinberg (2001), Pauling *et al.* (1949)). The resulting hemoglobin, HbS,
40 causes accelerated red cell destruction, erythroid hyperplasia and painful vaso-occlusive 'crises' (Steinberg (2001)). Vaso-occlusion can damage organs, eventually causing long-term disabilities (e.g. following stroke or bone necrosis), and sometimes sudden death. While a very serious disorder, the course of sickle cell disease is typically unpredictable (Steinberg (2001)). By increasing production of fetal hemoglobin (Swank and Stamatoyannopoulos (1998)) and suppressing hematopoiesis, hydroxyurea can produce a measurable clinical benefit (Platt *et al.* (1984)), Charache *et al.* (1992), Atweh and Loukopoulos (2001)). Since hydroxyurea is a cytotoxic agent, there is a great need for alternative,
45 less toxic drugs to induce γ -globin gene expression (Perrine *et al.* (2005), Stamatoyannopoulos (2005)). As for β -thalassemia, allogeneic bone marrow transplantation (BMT) is at present the only curative therapy for sickle cell disease (Tisdale and Sadelain (2001), Vermynen *et al.* (1998), Luzzatto and Goodfellow (1989)).

- [0006]** BMT, however, is not available as a therapeutic option to most patients suffering from β -thalassemia or sickle
50 cell disease, due to the lack of an HLA-matched bone marrow donor for most individuals. Furthermore, although potentially curative, allogeneic BMT is not devoid of complications. Safe transplantation requires the identification of a histo-compatible donor to minimize the risks of graft rejection and graft-versus-host disease (Tisdale and Sadelain (2001), Vermynen *et al.* (1998), Luzzatto and Goodfellow (1989)). Because of the greater risks associated with matched-unrelated or mismatched transplants, most patients have to settle for life-long transfusion therapy, which does not correct ineffective
55 erythropoiesis and exacerbates systemic iron accumulation. Moreover, despite the considerable improvement in life expectancy in the last decades (Borgna-Pignatti *et al.* (2004), Telfer *et al.* (2009), Ladis *et al.* (2011)), the risk of some serious complications arising over the long term from viral infections, iron toxicity and liver cirrhosis, remain (Mancuso *et al.* (2006)). These medical risks, together with the socio-economic cost of chronic β -thalassemia, underscore the need

for safe, effective and curative therapies.

[0007] The only means to cure rather than treat severe β -thalassemia is to provide the patient with healthy hematopoietic stem cells (HSCs). HSCs normally give rise to all blood cell types, including 20 billion RBCs per day in adults. HSCs can be harvested from a donor with wild-type β -globin genes to yield long-lived red blood cells (RBCs) with a normal content in hemoglobin. Alternatively, one may genetically correct the patient's own HSCs, which at once resolves the search for a donor and eliminates the risks of graft-versus-host disease and graft rejection associated with allogeneic BMT (Sadelain (1997), Sadelain *et al.* (2007)). Globin gene transfer aims to restore the capacity of the β -thalassemic subject's own blood-forming stem cells to generate RBCs with a sufficient hemoglobin content Sadelain *et al.* (2007), Persons and Tisdale (2004), Sadelain (2006)). The goal in patients with sickle cell anemia is to prevent sickling, which can be achieved by diluting the endogenous HbS with a non-sickling Hb that incorporates the vector-encoded globin chain. The patient's own HSCs are the cells that have to be genetically modified to ensure long-lasting therapeutic benefits and achieve a curative stem cell-based therapy.

[0008] The implementation of globin gene transfer for the treatment of severe β -thalassemia and sickle cell anemia requires the efficient introduction of a regulated human β - or β -like globin gene in HSCs. The β -globin gene (or β -like variant) must be expressed in erythroid-specific fashion and at high level, especially for the treatment of transfusion-dependent beta-zero thalassemias.

[0009] The globin vectors developed to date present shortcomings that may limit or even preclude their safe use in thalassemia and sickle cell patients. Some of the β -globin locus control region (LCR) components contained in the vectors, in particular Dnase I hypersensitive site-2 (HS2), may have non-erythroid activity, exposing patients to the risk of insertional oncogenesis as seen with non-specific expression vectors. Further, the use of large LCR segments can be detrimental to the production of high titer vectors and the efficient transduction of patients HSCs. Accordingly, there is a need for novel globin expression cassettes that allow for therapeutic expression of a globin gene (e.g., human β -globin gene) in erythroid-specific and differentiation stage-specific fashion with minimal risk of insertional oncogenesis, and that enable high level transduction, thus improving their safety when used in treating thalassemia and sickle cell patients.

SUMMARY OF THE INVENTION

[0010] The presently disclosed subject matter generally provides enhancer blocking insulators, and certain insulators additionally possess barrier insulator activity. The presently disclosed subject matter also provides expression cassettes comprising one or more insulators and allows for expression of a globin gene (e.g., a human β globin gene). Also provided are vectors comprising such expression cassettes, cells transduced with such expression cassettes or such vectors, and uses of such expression cassettes for treating hemoglobinopathies (e.g., β -thalassemia and sickle cell anemia).

[0011] In certain non-limiting embodiments, the presently disclosed subject matter provides an insulator comprising the CTCF binding site sequence set forth in SEQ ID NO:18, for example, but not limited to, an insulator comprising SEQ ID NO: 24 or SEQ ID NO:25, such as an insulator having the nucleotide sequence set forth in SEQ ID NO:1 (and see *infra*). The presently disclosed subject matter also provides expression cassettes comprising at least one insulator comprising the CTCF binding site sequence set forth in SEQ ID NO:18, for example, but not limited to, an insulator comprising SEQ ID NO: 24 or SEQ ID NO:25, such as an insulator having the nucleotide sequence set forth in SEQ ID NO:1. In a non-limiting embodiment, an expression cassette comprises at least one insulator comprising the CTCF binding site sequence set forth in SEQ ID NO:18, for example, but not limited to, an insulator comprising SEQ ID NO: 24 or SEQ ID NO:25, such as an insulator having the nucleotide sequence set forth in SEQ ID NO:1, and a globin gene or a functional portion thereof operably linked to a β -globin locus control region (LCR). In certain embodiments, the β -globin LCR does not comprise a Dnase I hypersensitive site-2 (HS2) region. In certain embodiments, the β -globin LCR does not comprise a core sequence of HS2. In one non-limiting embodiment, the core sequence of HS2 has the nucleotide sequence set forth in SEQ ID NO:20. In one non-limiting embodiment, the core sequence of HS2 has the nucleotide sequence set forth in SEQ ID NO:21. In certain embodiments, the β -globin LCR does not comprise a HS2 region that sustains the enhancer activity of HS2. In one non-limiting embodiment, the β -globin LCR comprises a Dnase I hypersensitive site-1 (HS1) region, a Dnase I hypersensitive site-3 (HS3) region, and a Dnase I hypersensitive site-4 (HS4) region. In certain embodiments, the HS3 region is positioned between the HS1 and the HS4 region.

[0012] In certain embodiments, the HS1 region is about 1.1 kb bp in length. In one non-limiting embodiment, the HS1 region is between about 500bp and about 1000 bp in length. In one non-limiting embodiment, the HS1 region has the nucleotide sequence set forth in SEQ ID NO:2. In certain embodiments, the HS1 region is about 600 bp in length. In one non-limiting embodiment, the HS1 region is 602 bp in length. In certain embodiments, the HS1 region is between about 500 and about 600 bp in length. In one non-limiting embodiment, the HS1 region has the nucleotide sequence set forth in SEQ ID NO:3. In certain embodiments, the HS1 region is about 490 bp in length. In one non-limiting embodiment, the HS1 region is 489 bp in length. In one non-limiting embodiment, the β -globin LCR comprises a HS1 region having a nucleotide

sequence set forth in SEQ ID NO:2, a HS3 region having a nucleotide sequence set forth in SEQ ID NO:5, and a HS4 region having a nucleotide sequence set forth in SEQ ID NO:6, and the β -globin LCR region does not comprise a HS2 region. In one non-limiting embodiment, the β -globin LCR region comprises a HS1 region having a nucleotide sequence set forth in SEQ ID NO:3, a HS3 region having a nucleotide sequence set forth in SEQ ID NO:5, and a HS4 region having a nucleotide sequence set forth in SEQ ID NO:8, and the β -globin LCR does not comprise a HS2 region. In one non-limiting embodiment, the β -globin LCR comprises a HS1 region having a nucleotide sequence set forth in SEQ ID NO:4, a HS3 region having a nucleotide sequence set forth in SEQ ID NO:5, and a HS4 region having a nucleotide sequence set forth in SEQ ID NO:8, and the β -globin LCR does not comprise a HS2 region.

[0013] In certain embodiments, the β -globin LCR region does not comprise a HS1 region and/or does not comprise a HS2 region, and the β -globin LCR does not comprise a core sequence of HS2. In certain embodiments, the β -globin LCR does not comprise a core sequence of HS1. In one non-limiting embodiment, the core sequence of HS1 has the nucleotide sequence set forth in SEQ ID NO:22. In one non-limiting embodiment, the core sequence of HS1 has the nucleotide sequence set forth in SEQ ID NO:23. In certain embodiments, the β -globin LCR does not comprise a HS1 region that sustains the function of HS1. In certain embodiments, the β -globin LCR comprises a HS3 region and a HS4 region and does not comprise a core sequence of HS1. In certain embodiments, the HS3 region is positioned between a globin gene or functional portion thereof and the HS4 region. In certain embodiments, the HS3 region is between about 200 and about 1400bp in length, e.g., between about 1300 and 1400 bp in length. In certain embodiments, the HS3 region is about 1300bp in length. In one non-limiting embodiment, the HS3 region is 1301bp in length. In one non-limiting embodiment, the HS3 region has the nucleotide sequence set forth in SEQ ID NO:5. In certain embodiments, the HS4 region is between about 200 and about 1200bp in length, e.g., between about 400 and 1100 bp in length. In certain embodiments, the HS4 region is about 1.1 kb in length. In one non-limiting embodiment, the HS4 region is 1065bp in length. In one non-limiting embodiment, the HS4 region has the nucleotide sequence set forth in SEQ ID NO:6. In one non-limiting embodiment, the HS4 region has the nucleotide sequence set forth in SEQ ID NO:7. In certain embodiments, the HS4 region is about 450 bp in length. In one non-limiting embodiment, the HS4 region is 446 bp in length. In one non-limiting embodiment, the HS4 region has the nucleotide sequence set forth in SEQ ID NO:8. In one non-limiting embodiment, the β -globin LCR region comprises a HS3 region having the nucleotide sequence set forth in SEQ ID NO:5 and a HS4 region having a nucleotide sequence set forth in SEQ ID NO:6, and the β -globin LCR region does not comprise a HS1 region or a HS2 region.

[0014] Alternatively, the β -globin LCR region can comprise a HS2 region, a HS3 region, and a HS4 region. In certain embodiments, the HS2 region is between about 400 and about 1000bp in length, e.g., between about 800 and 900 bp in length. In certain embodiments, the HS2 region is about 860bp in length. In one non-limiting embodiment, the HS2 region has the nucleotide sequence set forth in SEQ ID NO:9. In certain embodiments, the HS3 region is about 1300bp in length. In one non-limiting embodiment, the HS3 region is 1301bp in length. In one non-limiting embodiment, the HS3 region has the nucleotide sequence set forth in SEQ ID NO:5. In certain embodiments, the HS4 region is about 1.1 kb in length. In one non-limiting embodiment, the HS4 region is 1065bp in length. In one non-limiting embodiment, the HS4 region has the nucleotide sequence set forth in SEQ ID NO:7. In one non-limiting embodiment, the β -globin LCR region comprises a HS2 region having the nucleotide sequence set forth in SEQ ID NO:9, a HS3 region having the nucleotide sequence set forth in SEQ ID NO:5, and a HS4 region having the nucleotide sequence set forth in SEQ ID NO:7. Additionally, the β -globin LCR region can further comprise a HS1 region.

[0015] In certain embodiments, the globin gene is selected from the group consisting of β -globin gene, γ -globin gene, and δ -globin gene. In one non-limiting embodiment, the globin gene is human β -globin gene. In non-limiting embodiments, the human β -globin gene is selected from the group consisting of a wild-type human β -globin gene, a deleted human β -globin gene comprising one or more deletions of intron sequences, and a mutated human β -globin gene encoding at least one anti-sickling amino acid residue. In one non-limiting embodiment, the human β -globin gene is human β^A -globin gene encoding a threonine to glutamine mutation at codon 87 (β^A -T87Q).

[0016] In certain embodiments, the expression cassette comprises one insulator comprising the CTCF binding site sequence set forth in SEQ ID NO:18, for example, but not limited to, an insulator comprising SEQ ID NO: 24 or SEQ ID NO:25, such as an insulator having the nucleotide sequence set forth in SEQ ID NO:1. In certain embodiments, the expression cassette comprises two insulators, each comprising the CTCF binding site sequence set forth in SEQ ID NO:18, for example, but not limited to, where one or both insulators comprise SEQ ID NO: 24 or SEQ ID NO:25 and/or have the nucleotide sequence set forth in SEQ ID NO:1.

[0017] In certain embodiments, the expression cassette further comprises a β -globin promoter. In certain embodiments, the β -globin promoter is positioned between the globin gene or functional portion thereof and β -globin LCR region. In certain embodiments, the β -globin promoter is between about 200 and about 700bp in length. In one non-limiting embodiment, the β -globin promoter is a human β -globin promoter that is about 613bp in length. In one non-limiting embodiment, the human β -globin promoter has the nucleotide sequence set forth in SEQ ID NO:10. In another non-limiting embodiment, the β -globin promoter is a human β -globin promoter that is about 265bp in length. In one non-limiting embodiment, the β human β -globin promoter has the nucleotide sequence set forth in SEQ ID NO:11.

[0018] In certain embodiments, the expression cassette further comprises a human β -globin 3' enhancer. In certain embodiments, the human β -globin 3' enhancer is positioned in the upstream of the globin gene or functional portion thereof. In certain embodiments, the β -globin 3' enhancer is between about 700 and about 900bp in length, *e.g.*, between about 800 and 900 bp in length. In one non-limiting embodiment, the human β -globin 3' enhancer is about 879 bp in length. In one non-limiting embodiment, the human β -globin 3' enhancer has the nucleotide sequence set forth in SEQ ID NO:12.

[0019] In certain embodiments, the expression cassette further comprises at least one erythroid-specific enhancer. In certain embodiments, the at least one erythroid-specific enhancer is positioned between the globin gene or functional portion thereof and the β -globin LCR region. In certain embodiments, the at least one erythroid-specific enhancer has a nucleotide sequence selected from the group consisting of SEQ ID NOS: 13, 14, 15, 16 and 17. In certain embodiments, the at least one erythroid-specific enhancer is between about 100 and about 200bp in length. In certain embodiments, the expression cassette comprises one, two or three erythroid-specific enhancers.

[0020] In certain embodiments, the expression cassette allows for expression of the globin gene or functional portion thereof in a mammal. In one non-limiting embodiment, the expression cassette allows for expression of a human β -globin gene. In certain embodiments, the expression of the globin gene or functional portion thereof is restricted to erythroid tissue.

[0021] The presently disclosed subject matter also provides recombinant vectors comprising the above-described expression cassettes. In certain embodiments, the recombinant vector is a retroviral vector. In one non-limiting embodiment, the retroviral vector is a lentivirus vector. In certain embodiments, the expression cassette comprised in the recombinant vector comprises one insulator. In certain embodiments, the recombinant vector further comprises a Woodchuck hepatitis post-regulatory element (WPRE) in the 3' long terminal repeat (LTR) of the vector. In certain embodiments, the recombinant vector further comprises a bovine growth hormone polyadenylation signal in the 3' long terminal repeat (LTR) of the vector.

[0022] In addition, the presently disclosed subject matter provides non-naturally occurring or engineered nucleases comprising the above-described expression cassettes. In certain embodiments, the nuclease is selected from the group consisting of a non-naturally occurring or engineered zinc-finger nuclease (ZFN), a non-naturally occurring or engineered meganuclease, and a non-naturally occurring or engineered transcription activator-like effector nuclease (TALEN). In certain embodiments, the nuclease comprises a DNA binding domain and a nuclease cleavage domain. In certain embodiments, the nuclease binds to a genomic safe harbor site. In certain embodiments, the nuclease generates a double strand break (DSB) at the genomic safe harbor site. In certain embodiments, the expression cassette comprised in the nuclease comprises two of the insulator having the nucleotide sequence set forth in SEQ ID NO:1. In certain embodiments, the nuclease allows for targeted delivery of the expression cassette. The presently disclosed subject matter also provides polynucleotides encoding the above-described nucleases, and vectors comprising the polynucleotides. In one non-limiting embodiment, the vector is a lentiviral vector.

[0023] Furthermore, the presently disclosed subject matter provides non-naturally occurring or engineered CRISPR-Cas systems comprising the above-described expression cassettes. In certain embodiments, the CRISPR-Cas system comprises a CRISPR-Cas nuclease and single-guide RNA. In certain embodiments, the CRISPR-Cas system binds to a genomic safe harbor site. In certain embodiments, the CRISPR-Cas system generates a double strand break (DSB) at the genomic safe harbor site. In certain embodiments, the expression cassette comprised in the CRISPR-Cas system comprises two of the insulator having the nucleotide sequence set forth in SEQ ID NO:1. In certain embodiments, the CRISPR-Cas allows for targeted delivery of the expression cassette. The presently disclosed subject matter also provides polynucleotides encoding the above-described CRISPR-Cas systems, and vectors comprising the polynucleotides. In one non-limiting embodiment, the vector is a lentiviral vector.

[0024] In some embodiments, the genomic safe harbor site is an extragenic genomic safe harbor site. In certain embodiments, the genomic safe harbor site is located on chromosome 1. In some embodiments, the genomic safe harbor meets all of the following five criteria: (1) distance of at least 50 kb from the 5' end of any gene (*e.g.*, from the 5' end of the gene), (ii) distance of at least 300 kb from any cancer-related gene, (iii) within an open/accessible chromatin structure (measured by DNA cleavage with natural or engineered nucleases), (iv) location outside a gene transcription unit and (v) location outside ultraconserved regions (UCRs), microRNA or long non-coding RNA of the human genome.

[0025] Additionally, the presently disclosed subject matter provides cells transduced with the above-described expression cassettes, cells transduced with the above-described recombinant vectors, cells transduced with the above-described nucleases, cells transduced with the above-described CRISPR-Cas systems. In addition, the presently disclosed subject matter provides cells transduced with the above-described vectors. In certain embodiments, the cell is selected from the group consisting of a hematopoietic stem cell, an embryonic stem cell, an induced pluripotent stem cell, and a hemogenic endothelium cell. In one non-limiting embodiment, the hematopoietic stem cell is a CD34⁺ hematopoietic stem cell. In certain embodiments, the cell is transduced *ex vivo*.

[0026] Also provided are pharmaceutical compositions comprising an effective amount of the above-described cells and a pharmaceutically acceptable carrier. The presently disclosed subject matter also provides pharmaceutical com-

positions for treating a hemoglobinopathy comprising an effective amount of the above-described cells and a pharmaceutically acceptable carrier.

[0027] Furthermore, the presently disclosed subject matter provides kits for treating a hemoglobinopathy comprising the above-described cells. In certain embodiments, the kits further comprise written instructions for using the cell for treating a subject having a hemoglobinopathy.

[0028] In addition, the presently disclosed subject matter provides methods of treating a hemoglobinopathy in a subject, comprising administering an effective amount of the above-described cells to the subject, thereby restoring the subject's ability to produce red blood cells containing normal hemoglobin. In certain embodiments, a therapeutically relevant level of hemoglobin is produced in the subject following administering the cell to the subject. In certain amendments, the method comprises administering an effective amount of the cell transduced with the above-described recombinant vector. In some embodiments, the vector copy number of the recombinant vector in the cell that provides for the therapeutically relevant level of hemoglobin in the subject is about 0.5-2 vector copy number per cell. In certain embodiments, the method corrects ineffective erythropoiesis in the subject. In certain embodiments, the method does not incur the risk of graft-versus-host disease in the subject. In certain embodiments, the method does not comprise administering an immunosuppressive agent. In certain embodiments, the cell is selected from the group consisting of a hematopoietic stem cell, an embryonic stem cell, an induced pluripotent stem cell, and a hemogenic endothelium cell. In one non-limiting embodiment, the subject is a human. In certain embodiments, the cell is from the subject. In one non-limiting embodiment, the cell is from bone marrow of the subject.

[0029] In accordance with the presently disclosed subject matter, the hemoglobinopathy can be selected from the group consisting of hemoglobin C disease, hemoglobin sickle cell disease (SCD), sickle cell anemia, hereditary anemia, thalassemia, β -thalassemia, thalassemia major, thalassemia intermedia, α -thalassemia, and hemoglobin H disease. In one non-limiting embodiment, the hemoglobinopathy is β -thalassemia. In another non-limiting embodiment, the hemoglobinopathy is sickle cell anemia.

[0030] Aspects or embodiments of the invention may also be provided according to the following paragraphs:

1. An insulator comprising the CTCF binding site sequence set forth in SEQ ID NO:18.
2. The insulator of paragraph 1, comprising SEQ ID NO: 24 or SEQ ID NO:25
3. The insulator of paragraph 2, having the nucleotide sequence set forth in SEQ ID NO:1.
4. An expression cassette comprising the insulator of any of paragraphs 1-3, and a globin gene or a functional portion thereof operably linked to a β -globin locus control region (LCR) region.
5. The expression cassette of paragraph 4, wherein the β -globin LCR region does not comprise a Dnase I hypersensitive site-2 (HS2) region.
6. The expression cassette of paragraph 5, wherein the β -globin LCR region does not comprise a core sequence of HS2.
7. The expression cassette of paragraph 6, wherein the core sequence of HS2 has the nucleotide sequence set forth in SEQ ID NO:20.
8. The expression cassette of paragraph 6, wherein the core sequence of HS2 has the nucleotide sequence set forth in SEQ ID NO:21.
9. The expression cassette of paragraph 4, wherein the β -globin LCR region does not comprise a HS2 region that sustains the enhancer activity of HS2.
10. The expression cassette of any one of paragraphs 4-9, wherein the β -globin LCR region comprises a Dnase I hypersensitive site-1 (HS1) region, a Dnase I hypersensitive site-3 (HS3) region, and a Dnase I hypersensitive site-4 (HS4) region.
11. The expression cassette of paragraph 10, wherein the HS3 region is positioned between the HS1 region and the HS4 region.
12. The expression cassette of paragraph 10 or 11, wherein the HS1 region is about 1.1 kb in length.
13. The expression cassette of paragraph 12, wherein the HS1 region has the nucleotide sequence set forth in SEQ ID NO:2.
14. The expression cassette of paragraph 10 or 11, wherein the HS1 region is about 600 bp in length.
15. The expression cassette of paragraph 14, wherein the HS1 region is 602 bp in length.
16. The expression cassette of paragraph 15, wherein the HS1 region has the nucleotide sequence set forth in SEQ ID NO:3.
17. The expression cassette of paragraph 10 or 11, wherein the HS1 region is about 490 bp in length.
18. The expression cassette of paragraph 17, wherein the HS1 region is 489 bp in length.
19. The expression cassette of paragraph 18, wherein the HS1 region has the nucleotide sequence set forth in SEQ ID NO:4.
20. The expression cassette of paragraph 4 or 5, wherein the β -globin LCR region does not comprise a HS1 region.
21. The expression cassette of paragraph 20, wherein the β -globin LCR region does not comprise a core sequence

of HS1.

22. The expression cassette of paragraph 21, wherein the core sequence of HS1 has the nucleotide sequence set forth in SEQ ID NO:22.

23. The expression cassette of paragraph 21, wherein the core sequence of HS1 has the nucleotide sequence set forth in SEQ ID NO:23.

24. The expression cassette of paragraph 20, wherein the β -globin LCR region does not comprise a HS1 region that sustains the function of HS1.

25. The expression cassette of any one of paragraphs 20-24, wherein the β -globin LCR region comprises a HS3 region and a HS4 region.

26. The expression cassette of paragraph 25, wherein the HS3 region is positioned between the globin gene or functional portion thereof and the HS4 region.

27. The expression cassette of any one of paragraphs 10-26, wherein the HS3 region is about 1300 bp in length.

28. The expression cassette of paragraph 27, wherein the HS3 region is 1301 bp in length.

29. The expression cassette of paragraph 28, wherein the HS3 region has the nucleotide sequence set forth in SEQ ID NO:5.

30. The expression cassette of any one of paragraphs 10-29, wherein the HS4 region is about 1.1 kb in length.

31. The expression cassette of paragraph 30, wherein the HS4 region is 1065bp in length.

32. The expression cassette of paragraph 31, wherein the HS4 region has the nucleotide sequence set forth in SEQ ID NO:6.

33. The expression cassette of paragraph 31, wherein the HS4 region has the nucleotide sequence set forth in SEQ ID NO:7.

34. The expression cassette of any one of paragraphs 10-29, wherein the HS4 region is about 450 bp in length.

35. The expression cassette of paragraph 34, wherein the HS4 region is 446 bp in length.

36. The expression cassette of paragraph 35, wherein the HS4 region has the nucleotide sequence set forth in SEQ ID NO:8.

37. The expression cassette of any one of paragraphs 5-11, wherein the β -globin LCR region comprises a HS1 region having the nucleotide sequence set forth in SEQ ID NO:2, a HS3 region having the nucleotide sequence set forth in SEQ ID NO:5, and a HS4 region having the nucleotide sequence set forth in SEQ ID NO:6, and the β -globin LCR region does not comprise a HS2 region.

38. The expression cassette of any one of paragraphs 5-11, wherein the β -globin LCR region comprises a HS1 region having the nucleotide sequence set forth in SEQ ID NO:3, a HS3 region having the nucleotide sequence set forth in SEQ ID NO:5, and a HS4 region having the nucleotide sequence set forth in SEQ ID NO:8, and the β -globin LCR region does not comprise a HS2 region.

39. The expression cassette of any one of paragraphs 5-11, wherein the β -globin LCR region comprises a HS1 region having the nucleotide sequence set forth in SEQ ID NO:4, a HS3 region having the nucleotide sequence set forth in SEQ ID NO:5, and a HS4 region having the nucleotide sequence set forth in SEQ ID NO:8, and the β -globin LCR region does not comprise a HS2 region.

40. The expression cassette of any one of paragraphs 20-25, wherein the β -globin LCR region comprises a HS3 region having the nucleotide sequence set forth in SEQ ID NO:5 and a HS4 region having the nucleotide sequence set forth in SEQ ID NO:6, and the β -globin LCR region does not comprise a HS1 region or a HS2 region.

41. The expression cassette of paragraph 4, wherein the β -globin LCR region comprises a HS2 region, a HS3 region, and a HS4 region.

42. The expression cassette of paragraph 41, wherein the HS2 region is 860bp in length.

43. The expression cassette of paragraph 42, wherein the HS2 region has the nucleotide sequence set forth in SEQ ID NO:9.

44. The expression cassette of any one of paragraphs 41-42, wherein the HS3 region is about 1300 bp in length.

45. The expression cassette of paragraph 44, wherein the HS3 region is 1301 bp in length.

46. The expression cassette of paragraph 45, wherein the HS3 region has the nucleotide sequence set forth in SEQ ID NO:5.

47. The expression cassette of any one of paragraphs 41-46, wherein the HS4 region is about 1.1 kb in length.

48. The expression cassette of paragraph 47, wherein the HS4 region is 1065 bp in length.

49. The expression cassette of paragraph 49, wherein the HS4 region has the nucleotide sequence set forth in SEQ ID NO:7.

50. The expression cassette of paragraph 41, wherein the β -globin LCR region comprises a HS2 region having the nucleotide sequence set forth in SEQ ID NO:9, a HS3 region having the nucleotide sequence set forth in SEQ ID NO:5, and a HS4 region having the nucleotide sequence set forth in SEQ ID NO:7.

51. The expression cassette of any one of paragraphs 41-50, wherein the β -globin LCR region further comprises a HS1 region.

52. The expression cassette of any one of paragraphs 4-51, wherein the globin gene is selected from the group consisting of β -globin gene, γ -globin gene, and δ -globin gene.

53. The expression cassette of paragraph 52, wherein the globin gene is human β -globin gene.

54. The expression cassette of paragraph 53, wherein the human β -globin gene is selected from the group consisting of a wild-type human β -globin gene, a deleted human β -globin gene comprising one or more deletions of intron sequences, and a mutated human β -globin gene encoding at least one anti-sickling amino acid residue.

55. The expression cassette of paragraph 54, wherein the human β -globin gene is human β^A -globin gene encoding a threonine to glutamine mutation at codon 87 (β^{A-T87Q}).

56. The expression cassette of any one of paragraphs 4-55, comprising one insulator having the nucleotide sequence set forth in SEQ ID NO:1.

57. The expression cassette of any one of paragraphs 4-55, comprising two of the insulator having the nucleotide sequence set forth in SEQ ID NO:1.

58. The expression cassette of any one of paragraphs 4-57, further comprising a β -globin promoter.

59. The expression cassette of paragraph 58, wherein the β -globin promoter is positioned between the globin gene or functional portion thereof and the β -globin LCR region.

60. The expression cassette of paragraph 58 or 59, wherein the β -globin promoter is a human β -globin promoter that is about 613bp in length.

61. The expression cassette of paragraph 60, wherein the human β -globin promoter has the nucleotide sequence set forth in SEQ ID NO:10.

62. The expression cassette of paragraph 58 or 59, wherein the β -globin promoter is a human β -globin promoter that is about 265bp in length.

63. The expression cassette of paragraph 62, wherein the human β -globin promoter has the nucleotide sequence set forth in SEQ ID NO:11.

64. The expression cassette of any one of paragraphs 4-63, further comprising a human β -globin 3' enhancer.

65. The expression cassette of paragraph 64, wherein the human β -globin 3' enhancer is positioned in the upstream of the globin gene or functional portion thereof.

66. The expression cassette of paragraph 64 or 65, wherein the human β -globin 3' enhancer is about 879 bp in length.

67. The expression cassette of paragraph 66, wherein the human β -globin 3' enhancer has the nucleotide sequence set forth in SEQ ID NO:12.

68. The expression cassette of any one of paragraphs 5-67, further comprising at least one erythroid-specific enhancer.

69. The expression cassette of paragraph 68, wherein the at least one erythroid-specific enhancer is positioned between the globin gene or functional portion thereof and β -globin LCR region.

70. The expression cassette of paragraph 68 or 69, wherein the at least one erythroid-specific enhancer has a nucleotide sequence selected from the group consisting of SEQ ID NOS: 13, 14, 15, 16 and 17.

71. The expression cassette of any one of paragraphs 68-70, comprising one, two or three erythroid-specific enhancers.

72. The expression cassette of any one of paragraphs 4-71, wherein the expression cassette allows for expression of the globin gene or functional portion thereof in a mammal.

73. The expression cassette of paragraph 72, wherein the expression cassette allows for expression of human β -globin gene.

74. The expression cassette of paragraph 71 or 72, wherein the expression of the globin gene or functional portion thereof is restricted to erythroid tissue.

75. A recombinant vector comprising the expression cassette of any one of paragraphs 4-74.

76. The recombinant vector of paragraph 75, wherein the recombinant vector is a retroviral vector.

77. The recombinant vector of paragraph 76, wherein the retroviral vector is a lentivirus vector.

78. The recombinant vector of any one of paragraphs 75-77, wherein the expression cassette comprises one insulator having the nucleotide sequence set forth in SEQ ID NO:1.

79. The recombinant vector of any one of paragraphs 75-77, further comprising a Woodchuck hepatitis post-regulatory element (WPRE) in the 3' long terminal repeat (LTR) of the vector.

80. The recombinant vector of paragraph 79, further comprising a bovine growth hormone polyadenylation signal in the 3' long terminal repeat (LTR) of the vector.

81. A non-naturally occurring or engineered nuclease comprising the expression cassette of any one of paragraphs 4-74.

82. The nuclease of paragraph 81, wherein the nuclease is selected from the group consisting of a non-naturally occurring or engineered zinc-finger nuclease (ZFN), a non-naturally occurring or engineered meganuclease, and a non-naturally occurring or engineered transcription activator-like effector nuclease (TALEN).

83. The nuclease of paragraph 81 or 82, wherein the nuclease comprises a DNA binding domain and a nuclease

cleavage domain.

84. The nuclease of any one of paragraphs 81-83, wherein the nuclease binds to a genomic safe harbor site.

85. The nuclease of paragraph 84, wherein the nuclease generates a double strand break (DSB) at the genomic safe harbor site.

86. The nuclease of paragraph 84 or 85, wherein the genomic safe harbor site is an extragenic genomic safe harbor site.

87. The nuclease of any one of paragraphs 84-86, wherein the genomic safe harbor site is located on chromosome 1.

88. The nuclease of any one of paragraphs 84-86, wherein the genomic safe harbor site meets all of the following five criteria: (i) distance of at least 50 kb from the 5' end of any gene (e.g., from the 5' end of the gene), (ii) distance of at least 300 kb from any cancer-related gene, (iii) within an open/accessible chromatin structure (measured by DNA cleavage with natural or engineered nucleases), (iv) location outside a gene transcription unit and (v) location outside ultraconserved regions (UCRs), microRNA or long non-coding RNA of the human genome.

89. The nuclease of any one of paragraphs 81-88, wherein the expression cassette comprises two of the insulator having the nucleotide sequence set forth in SEQ ID NO:1.

90. The nuclease of any one of paragraphs 81-89, which allows for targeted delivery of the expression cassette.

91. A polynucleotide encoding the nuclease of any one of paragraphs 81-90.

92. A vector comprising the polynucleotide of paragraph 91.

93. The vector of paragraph 92, wherein the vector is a lentiviral vector.

94. A non-naturally occurring or engineered CRISPR-Cas system comprising the expression cassette of any one of paragraphs 4-74.

95. The system of paragraph 94, wherein the CRISPR-Cas system comprises a CRISPR-Cas nuclease and a single-guide RNA.

96. The system of paragraph 94 or 95, wherein the CRISPR-Cas system binds to a genomic safe harbor site.

97. The system of paragraph 96, wherein the CRISPR-Cas system generates a double strand break (DSB) at the genomic safe harbor site.

98. The system of paragraph 95 or 96, wherein the genomic safe harbor site is an extragenic genomic safe harbor site.

99. The system of any one of paragraphs 95-98, wherein the genomic safe harbor site is located on chromosome 1.

100. The system of any one of paragraphs 95-99, wherein the genomic safe harbor site meets all of the following five criteria: (i) distance of at least 50 kb from the 5' end of any gene (e.g., from the 5' end of the gene), (ii) distance of at least 300 kb from any cancer-related gene, (iii) within an open/accessible chromatin structure (measured by DNA cleavage with natural or engineered nucleases), (iv) location outside a gene transcription unit and (v) location outside ultraconserved regions (UCRs), microRNA or long non-coding RNA of the human genome.

101. The system of any one of paragraphs 94-100, wherein the expression cassette comprises two of the insulator having the nucleotide sequence set forth in SEQ ID NO:1.

102. The system of any one of paragraphs 94-101, which allows for targeted delivery of the expression cassette.

103. A polynucleotide encoding the CRISPR-Cas system of any one of paragraphs 94-102.

104. A vector comprising the polynucleotide of paragraph 103.

105. The vector of paragraph 104, wherein the vector is a lentiviral vector.

106. A cell transduced with the expression cassette of any one of paragraphs 4-74.

107. A cell transduced with the recombinant vector of any one of paragraphs 75-80.

108. A cell transduced with the nuclease of any one of paragraphs 81-90.

109. A cell transduced with the CRISPR-Cas system of any one of paragraphs 94-102.

110. A cell transduced with the vector of any one of paragraphs 92, 93, 104 and 105.

111. The cell of any one of paragraphs 106-110, wherein the cell is selected from the group consisting of a hematopoietic stem cell, an embryonic stem cell, an induced pluripotent stem cell, and a hemogenic endothelium cell.

112. The cell of paragraph 111, wherein the hematopoietic stem cell is a CD34⁺ hematopoietic stem cell.

113. The cell of any one of paragraphs 106-112, wherein the cell is transduced *ex vivo*.

114. A pharmaceutical composition comprising an effective amount of the cell of any one of paragraphs 106-113 and a pharmaceutically acceptable carrier.

115. A pharmaceutical composition for treating a hemoglobinopathy comprising an effective amount of the cell of any one of paragraphs 106-113 and a pharmaceutically acceptable carrier.

116. The pharmaceutical composition of paragraph 115, wherein the hemoglobinopathy is selected from the group consisting of hemoglobin C disease, hemoglobin sickle cell disease (SCD), sickle cell anemia, hereditary anemia, thalassemia, β -thalassemia, thalassemia major, thalassemia intermedia, α -thalassemia, and hemoglobin H disease.

117. The pharmaceutical composition of paragraph 116, wherein the hemoglobinopathy is β -thalassemia.

118. The pharmaceutical composition of paragraph 116, wherein the hemoglobinopathy is sickle cell anemia.

119. A kit for treating a hemoglobinopathy comprising the cell of any one of paragraphs 106-113.

120. The kit of paragraph 119, wherein the kit further comprises written instructions for using the cell for treating a

subject having a hemoglobinopathy.

121. The kit of paragraph 119 or 120, wherein the hemoglobinopathy is selected from the group consisting of hemoglobin C disease, hemoglobin sickle cell disease (SCD), sickle cell anemia, hereditary anemia, thalassemia, β -thalassemia, thalassemia major, thalassemia intermedia, α -thalassemia, and hemoglobin H disease.

122. The kit of paragraph 121, wherein the hemoglobinopathy is β -thalassemia.

123. The kit of paragraph 121, wherein the hemoglobinopathy is sickle cell anemia.

124. A method of treating a hemoglobinopathy in a subject, comprising administering an effective amount of the cell of any one of paragraphs 106-113 to the subject, thereby restoring the subject's ability to produce red blood cells containing normal hemoglobin.

125. The method of paragraph 124, wherein a therapeutically relevant level of hemoglobin is produced in the subject following administering the cell to the subject.

126. A method comprising administering an effective amount of the cell transduced with the recombinant vector of any one of paragraphs 75-80.

127. The method of paragraph 126, wherein the vector copy number of the recombinant vector in the cell that provides for the therapeutically relevant level of hemoglobin in the subject is about 0.5-2 vector copy number per cell.

128. The method of any one of paragraphs 124-127, wherein the method corrects ineffective erythropoiesis in the subject.

129. The method of any one of paragraphs 124-128, wherein the method does not incur the risk of graft-versus-host disease to the subject.

130. The method of any one of paragraphs 124-129, wherein the method does not comprise administering an immunosuppressive agent.

131. The method of any one of paragraphs 124-130, wherein the hemoglobinopathy is selected from the group consisting of hemoglobin C disease, hemoglobin sickle cell disease (SCD), sickle cell anemia, hereditary anemia, thalassemia, β -thalassemia, thalassemia major, thalassemia intermedia, α -thalassemia, and hemoglobin H disease.

132. The method of paragraph 131, wherein the hemoglobinopathy is β -thalassemia.

133. The method of paragraph 131, wherein the hemoglobinopathy is sickle cell anemia.

134. The method of any one of paragraphs 124-133, wherein the cell is selected from the group consisting of a hematopoietic stem cell, an embryonic stem cell, an induced pluripotent stem cell, and a hemogenic endothelium cell.

135. The method of any one of paragraphs 124-134, wherein the subject is a human.

136. The method of any one of paragraphs 124-135, wherein the cell is from the subject.

137. The method of paragraph 136, wherein the cell is from bone marrow of the subject.

BRIEF DESCRIPTION OF THE FIGURES

[0031] The following Detailed Description, given by way of example, but not intended to limit the invention to specific embodiments described, may be understood in conjunction with the accompanying drawings.

Figure 1 depicts a recombinant vector comprising an expression cassette in accordance with one non-limiting embodiment of the presently disclosed subject matter.

Figure 2 depicts a recombinant vector an expression cassette in accordance with one non-limiting embodiment of the presently disclosed subject matter.

Figure 3 depicts a recombinant vector an expression cassette in accordance with one non-limiting embodiment of the presently disclosed subject matter.

Figure 4 depicts a recombinant vector an expression cassette in accordance with one non-limiting embodiment of the presently disclosed subject matter.

Figures 5A-C represent the genotoxicity of insulator A1. (A) demonstrate the gammaretroviral vector genotoxicity assay used. (B) notice the increased survival of mice receiving 32D cells transduced with insulated gammaretroviral vector. Also notice the results obtained with cHS4 and with the uninsulated control. (C) show that insulator A1 decreased the risk of genotoxicity.

Figure 6 represents normalized β chain expression in thalassemic Hbb^{th3/+} mice 8 and 44 weeks post-treatment.

Figure 7 represents the evaluation of enhancer activity in non-erythroid K562 cells.

Figure 8 represents the erythroid-specific enhancers in accordance with certain embodiments of the presently disclosed subject matter.

Figure 9 represents the erythroid-specific enhancers in accordance with certain embodiments of the presently disclosed subject matter.

Figures 10A-B depict various recombinant vectors comprising the presently disclosed expression cassettes.

Figure 11 represents the titer of the recombinant vectors comprising the presently disclosed expression cassettes.

Figure 12 represents the titer of the recombinant vectors comprising the presently disclosed expression cassettes.

DETAILED DESCRIPTION OF THE INVENTION

[0032] The presently disclosed subject matter generally provides expression cassettes that allow for expression of a globin gene (e.g., human β -globin gene). In one non-limiting example, the expression cassette comprises at least one insulator comprising the CTCF binding site sequence set forth in SEQ ID NO:18, for example, but not limited to, an insulator comprising SEQ ID NO: 24 or SEQ ID NO:25, such as an insulator having the nucleotide sequence set forth in SEQ ID NO:1 and a globin gene or a functional portion thereof operably linked to a β -globin locus control region (LCR) region. The expression of the globin gene induced by the presently disclosed expression cassettes is erythroid-specific, differentiation stage-specific, high-level, and sustained. The presently disclosed subject matter also provides recombinant vectors, non-naturally occurring or engineered nucleases, and non-naturally occurring or engineered CRISPR-Cas systems comprising such expression cassettes, and cells transduced with such expression cassettes, recombinant vectors, nucleases and CRISPR-Cas systems. The presently disclosed expression cassettes and vectors comprising thereof provide for a safe gene transfer therapy as therapeutic transgene expression is achieved (e.g., a therapeutically relevant level of hemoglobin is produced) with a low vector copy number per cell (e.g., 0.5-2, 1-2, or even 0.5-1). In addition, the presently disclosed subject matter provides methods of using such transduced cells for treating a hemoglobinopathy (e.g., β -thalassemia and sickle cell anemia).

I. Definitions

[0033] Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. The following references provide one of skill with a general definition of many of the terms used in this invention: Singleton et al., Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger et al. (eds.), Springer Verlag (1991); and Hale & Marham, The Harper Collins Dictionary of Biology (1991). As used herein, the following terms have the meanings ascribed to them below, unless specified otherwise.

[0034] As used herein, the term "expression cassette" refers to a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements, which permit transcription of a particular nucleic acid in a target cell. The expression cassette can be incorporated into a plasmid, chromosome, mitochondrial DNA, plastid DNA, virus or nucleic acid region. The expression cassette portion can include a gene to be transcribed and elements that control the expression of the gene (e.g., a promoter).

[0035] As used herein, the term " β -globin locus control region (LCR) region" refers to a polynucleotide composed of one or more Dnase I hypersensitive site (HS) regions, including a HS1 region, a HS2 region, a HS3 region, and a HS4 region. The structure of many LCRs of the β -globin genes have been published, e.g., human (Li et al., J. Biol. Chem. (1985); 260:14,901; Li et al., Proc. Natl. Acad. Sci. (1990) 87:8207); mouse (Shehee et al., J. Mol. Biol. (1989);205:41); rabbit (Margot et al., J. Mol. Biol. (1989);205:15); and goat (Li, Q., et al., Genomics (1991);9:488), each of which are incorporated by reference herein. In certain embodiments, the β -globin LCR region comprises a HS2 region (e.g., a β -globin LCR region comprising a HS2 region, a HS3 region and a HS4 region; and a β -globin LCR region comprising a HS1 region, a HS2 region, a HS3 region and a HS4 region). In certain embodiments, the β -globin LCR region does not comprise a HS2 region (e.g., a β -globin LCR region comprising a HS1 region, a HS3 region, a HS4 region). In certain embodiments, the β -globin LCR region does not comprise a HS2 region or a HS1 region (e.g., a β -globin LCR region comprising a HS3 region and a HS4 region).

[0036] As used herein, the term "recombinant" includes reference to a cell or vector, that has been modified by the introduction of a heterologous nucleic acid or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found in identical form within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all as a result of deliberate human intervention or may have reduced or eliminated expression of a native gene.

[0037] As used herein, the term "globin" refers to a family of heme-containing proteins that are involved in the binding and transport of oxygen. Subunits of vertebrate and invertebrate hemoglobins, vertebrate and invertebrate myoglobins or mutants thereof are included by the term globin.

[0038] As used herein, the term "wild-type" refers to the normal gene, virus, or organism found in nature without any mutation or modification.

[0039] The terms "polynucleotide", "nucleotide", "nucleotide sequence", "nucleic acid" and "oligonucleotide" are used interchangeably. They refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three dimensional structure, and may perform any function, known or unknown. The following are non limiting examples of polynucleotides: coding or non-coding regions of a gene or gene region, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, short interfering RNA (siRNA), short-hairpin RNA (shRNA), micro-RNA (miRNA), ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated

RNA of any sequence, nucleic acid probes, and primers. A polynucleotide can comprise one or more modified nucleotides, such as methylated nucleotides and nucleotide analogs. In particular embodiments, the presently disclosed subject matter provides polynucleotides encoding one or more globin genes or functional portions thereof. If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component. Such polynucleotides need not be 100% identical with an endogenous nucleic acid sequence, but will typically exhibit substantial identity. Polynucleotides having "substantial identity" to an endogenous sequence are typically capable of hybridizing with at least one strand of a double-stranded nucleic acid molecule. By "hybridize" is meant pair to form a double-stranded molecule between complementary polynucleotide sequences (e.g., a gene described herein), or portions thereof, under various conditions of stringency. (See, e.g., Wahl, G. M. and S. L. Berger (1987) *Methods Enzymol.* 152:399; Kimmel, A. R. (1987) *Methods Enzymol.* 152:507).

[0040] For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and more preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and more preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30° C, more preferably of at least about 37° C, and most preferably of at least about 42° C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization will occur at 30° C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37° C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 µg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42° C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50% formamide, and 200 µg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

[0041] For most applications, washing steps that follow hybridization will also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include a temperature of at least about 25° C, more preferably of at least about 42° C, and even more preferably of at least about 68° C. In a preferred embodiment, wash steps will occur at 25° C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42° C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 68° C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art. Hybridization techniques are well known to those skilled in the art and are described, for example, in Benton and Davis (*Science* 196:180, 1977); Grunstein and Rogness (*Proc. Natl. Acad. Sci., USA* 72:3961, 1975); Ausubel et al. (*Current Protocols in Molecular Biology*, Wiley Interscience, New York, 2001); Berger and Kimmel (*Guide to Molecular Cloning Techniques*, 1987, Academic Press, New York); and Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York.

[0042] As used herein, the terms "polypeptide" and "protein" are used interchangeably to refer to a polymer of amino acid residues and to variants and synthetic analogues of the same. Thus, these terms apply to amino acid polymers in which one or more amino acid residues are synthetic non-naturally occurring amino acids, such as a chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally-occurring amino acid polymers. Particular embodiments of the presently disclosed subject matter also include polypeptide "variants." Polypeptide "variant" refers to polypeptides that are distinguished from a reference polypeptide by the addition, deletion, truncations, and/or substitution of at least one amino acid residue, and that retain a biological activity. In certain embodiments, a polypeptide variant is distinguished from a reference polypeptide by one or more substitutions, which may be conservative or non-conservative, as known in the art. In certain embodiments, a variant polypeptide includes an amino acid sequence having at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity or similarity to a corresponding sequence of a reference polypeptide. In certain embodiments, the amino acid additions or deletions occur at the C-terminal end and/or the N-terminal end of the reference polypeptide. In certain embodiments, the amino acid deletions include C-terminal truncations of about 1, about 2, about 3, about 4, about 5, about 6, about 7, about 8, about 9, about 10, about 15, about 20, about 25, about 30, about 35, about 40, about 45, 50, about 55, about 60, about 65, about 70, about 75, about 80, about 85, about 90, about 95, about 100, about 105, about 110, about 115, about 120, about 125, about 130, about 135, about 140, about 145, about 150, about 155, about 160, about 165, about 170, or about 175 or more amino acids, including all intervening numbers of amino acids, e.g., 25, 26, 27, 29, 30 ... 100, 101, 102, 103, 104, 105 ... 170, 171, 172, 173, 174, etc.

[0043] As noted above, polypeptides of the presently disclosed subject matter may be altered in various ways including amino acid substitutions, deletions, truncations, and insertions. Methods for such manipulations are generally known in the art. For example, amino acid sequence variants of a reference polypeptide can be prepared by mutations in the DNA. Methods for mutagenesis and nucleotide sequence alterations are well known in the art. See, for example, Kunkel (1985, Proc. Natl. Acad. Sci. USA. 82: 488-492), Kunkel et al., (1987, Methods in Enzymol, 154: 367-382), U.S. Pat. No. 4,873,192, Watson, J. D. et al., Molecular Biology of the Gene, Fourth Edition, Benjamin/Cummings, Menlo Park, Calif, 1987) and the references cited therein. Guidance as to appropriate amino acid substitutions that do not affect biological activity of the protein of interest may be found in the model of Dayhoff et al., (1978) Atlas of Protein Sequence and Structure (Natl. Biomed. Res. Found., Washington, D.C.).

[0044] As used herein, the term "substantially identical" refers to a polypeptide or a polynucleotide exhibiting at least 50% identity to a reference amino acid sequence (for example, any one of the amino acid sequences described herein) or a nucleic acid sequence (for example, any one of the nucleic acid sequences described herein). Preferably, such a sequence is at least 60%, more preferably 80% or 85%, and more preferably 90%, 95% or even 99% identical at the amino acid level or nucleic acid to the sequence used for comparison.

[0045] Sequence identity or homology is typically measured using sequence analysis software (for example, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705, BLAST, BESTFIT, GAP, or PILEUP/PRETTYBOX programs). Such software matches identical or similar sequences by assigning degrees of homology to various substitutions, deletions, and/or other modifications. In an exemplary approach to determining the degree of identity or homology, a BLAST program may be used, with a probability score between e-3 and e-100 indicating a closely related sequence. The percentage of identity between two sequences can also be determined with programs such as DNAMAN (Lynnon Biosoft, version 3.2). Using this program two sequences can be aligned using the optimal alignment algorithm (Smith and Waterman, 1981). After alignment of the two sequences the percentage identity can be calculated by dividing the number of identical nucleotides between the two sequences by the length of the aligned sequences minus the length of all gaps.

[0046] Terms that describe the orientation of polynucleotides include: 5' (normally the end of the polynucleotide having a free phosphate group) and 3' (normally the end of the polynucleotide having a free hydroxyl (OH) group). Polynucleotide sequences can be annotated in the 5' to 3' orientation or the 3' to 5' orientation.

[0047] As used herein, a "single guide RNA" or a "synthetic guide RNA" refers to the polynucleotide sequence comprising the guide sequence, the tracr sequence and the tracr mate sequence. The term "guide sequence" refers to the about 20 bp sequence within the guide RNA that specifies the target site and may be used interchangeably with the terms "guide" or "spacer". The term "tracr mate sequence" may also be used interchangeably with the term "direct repeat(s)".

[0048] The terms "non-naturally occurring" or "engineered" are used interchangeably and indicate the involvement of the hand of man. The terms, when referring to nucleic acid molecules or polypeptides mean that the nucleic acid molecule or the polypeptide is at least substantially free from at least one other component with which they are naturally associated in nature and as found in nature.

[0049] As used herein, the term "expression" refers to the process by which a polynucleotide is transcribed from a DNA template (such as into and mRNA or other RNA transcript) and/or the process by which a transcribed mRNA is subsequently translated into peptides, polypeptides, or proteins. Transcripts and encoded polypeptides may be collectively referred to as "gene product." If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell.

[0050] As used herein, the term "treating" or "treatment" refers to clinical intervention in an attempt to alter the disease course of the individual or cell being treated, and can be performed either for prophylaxis or during the course of clinical pathology. Therapeutic effects of treatment include, without limitation, preventing occurrence or recurrence of disease, alleviation of symptoms, diminishment of any direct or indirect pathological consequences of the disease, preventing metastases, decreasing the rate of disease progression, amelioration or palliation of the disease state, and remission or improved prognosis. By preventing progression of a disease or disorder, a treatment can prevent deterioration due to a disorder in an affected or diagnosed subject or a subject suspected of having the disorder, but also a treatment may prevent the onset of the disorder or a symptom of the disorder in a subject at risk for the disorder or suspected of having the disorder.

[0051] As used herein, the term "subject" refers to any animal (e.g., a mammal), including, but not limited to, humans, non-human primates, rodents, and the like (e.g., which is to be the recipient of a particular treatment, or from whom cells are harvested).

[0052] As used herein, the term "isolated cell" refers to a cell that is separated from the molecular and/or cellular components that naturally accompany the cell. As used herein, the term "isolated" refers to material that is free, substantially free, or essentially free to varying degrees from components which normally accompany it as found in its native state. "Isolate" denotes a degree of separation from original source or surroundings.

[0053] As used herein, the term "cell population" refers to a group of at least two cells expressing similar or different

phenotypes. In non-limiting examples, a cell population can include at least about 10, at least about 100, at least about 200, at least about 300, at least about 400, at least about 500, at least about 600, at least about 700, at least about 800, at least about 900, at least about 10^3 cells, at least about 10^4 cells, at least about 10^5 cells, at least about 10^6 cells, at least about 10^7 cells, or at least about 10^8 cells expressing similar or different phenotypes.

[0054] As used herein, the term "cleavage" refers to the breakage of the covalent backbone of a DNA molecule. Cleavage can be initiated by a variety of methods including, but not limited to, enzymatic or chemical hydrolysis of a phosphodiester bond. Both single-stranded cleavage and double-stranded cleavage are possible, and double-stranded cleavage can occur as a result of two distinct single-stranded cleavage events. DNA cleavage can result in the production of either blunt ends or staggered ends. In certain embodiments, fusion polypeptides are used for targeted double-stranded DNA cleavage.

[0055] As used herein, the term "cleavage half-domain" refers to a polypeptide sequence which, in conjunction with a second polypeptide (either identical or different) forms a complex having cleavage activity (preferably double-strand cleavage activity). The terms "first and second cleavage half-domains," "+ and - cleavage half-domains" and "right and left cleavage half-domains" are used interchangeably to refer to pairs of cleavage half-domains that dimerize.

[0056] As used herein, the term "chromosome" refers to a chromatin complex comprising all or a portion of the genome of a cell. The genome of a cell is often characterized by its karyotype, which is the collection of all the chromosomes that comprise the genome of the cell. The genome of a cell can comprise one or more chromosomes.

[0057] As used herein, the term "gene" includes a DNA region encoding a gene product, as well as all DNA regions which regulate the production of the gene product, whether or not such regulatory sequences are adjacent to coding and/or transcribed sequences. Accordingly, a gene includes, but is not limited to, promoter sequences, terminators, translational regulatory sequences such as ribosome binding sites and internal ribosome entry sites, enhancers, silencers, insulators, boundary elements, replication origins, matrix attachment sites and locus control regions.

[0058] The terms "operative linkage" and "operatively linked" (or "operably linked") are used interchangeably with reference to a juxtaposition of two or more components (such as sequence elements), in which the components are arranged such that both components function normally and allow the possibility that at least one of the components can mediate a function that is exerted upon at least one of the other components. By way of illustration, a transcriptional regulatory sequence, such as a promoter, is operatively linked to a coding sequence if the transcriptional regulatory sequence controls the level of transcription of the coding sequence in response to the presence or absence of one or more transcriptional regulatory factors. A transcriptional regulatory sequence is generally operatively linked in cis with a coding sequence, but need not be directly adjacent to it. For example, an enhancer is a transcriptional regulatory sequence that is operatively linked to a coding sequence, even though they are not contiguous.

[0059] A "functional region" or "functional portion" of a protein, polypeptide or nucleic acid is a protein, polypeptide or nucleic acid whose sequence is not identical to the full-length protein, polypeptide or nucleic acid, yet retains the same function as the full-length protein, polypeptide or nucleic acid. A functional region can possess more, fewer, or the same number of residues as the corresponding native molecule, and/or can contain one or more amino acid or nucleotide substitutions. Methods for determining the function of a nucleic acid (e.g., coding function, ability to hybridize to another nucleic acid) are well-known in the art. Similarly, methods for determining protein function are well-known. For example, the DNA-binding function of a polypeptide can be determined, for example, by filter-binding, electrophoretic mobility-shift, or immunoprecipitation assays. DNA cleavage can be assayed by gel electrophoresis. The ability of a protein to interact with another protein can be determined, for example, by co-immunoprecipitation, two-hybrid assays or complementation, both genetic and biochemical.

[0060] As used herein, the term "promoter" refers to a recognition site of a polynucleotide (DNA or RNA) to which an RNA polymerase binds. The term "enhancer" refers to a segment of DNA which contains sequences capable of providing enhanced transcription and in some instances can function independent of their orientation relative to another control sequence. An enhancer can function cooperatively or additively with promoters and/or other enhancer elements.

[0061] As used herein, the term "vector" refers to any genetic element, such as a plasmid, phage, transposon, cosmid, chromosome, virus, virion, etc., which is capable of replication when associated with the proper control elements and which can transfer gene sequences into cells. Thus, the term includes cloning and expression vehicles, as well as viral vectors and plasmid vectors.

[0062] As used herein, the term "modulate" refers to altering positively or negatively. Exemplary modulations include an about 1%, about 2%, about 5%, about 10%, about 25%, about 50%, about 75%, or about 100% change.

[0063] As used herein, the term "increase" refers to alter positively by at least about 5%, including, but not limited to, alter positively by about 5%, by about 10%, by about 25%, by about 30%, by about 50%, by about 75%, or by about 100%.

[0064] As used herein, the term "reduce" refers to alter negatively by at least about 5% including, but not limited to, alter negatively by about 5%, by about 10%, by about 25%, by about 30%, by about 50%, by about 75%, or by about 100%.

[0065] As used herein, the term "about" or "approximately" means within an acceptable error range for the particular value as determined by one of ordinary skill in the art, which will depend in part on how the value is measured or determined, i.e., the limitations of the measurement system. For example, "about" can mean within 3 or more than 3

standard deviations, per the practice in the art. Alternatively, "about" can mean a range of up to 20%, preferably up to 10%, more preferably up to 5%, and more preferably still up to 1% of a given value. Alternatively, particularly with respect to biological systems or processes, the term can mean within an order of magnitude, preferably within 5-fold, and more preferably within 2-fold, of a value.

II. Insulators

[0066] Several cases of vector-related malignant transformation have been reported in clinical settings, associated with the activation of cellular oncogenes by vector-encoded enhancers (Baum *et al.* (2006), Nienhuis *et al.* (2006), Ramezani *et al.* (2006)) and various vector modifications have been performed or proposed to reduce vector genotoxicity (Baum *et al.* (2006), Nienhuis *et al.* (2006), Ramezani *et al.* (2006)). A class of DNA elements known as chromatin insulators has been recognized as one approach to improve vector safety and performance (Emery (2011)).

[0067] Insulators are naturally occurring DNA elements that help from the functional boundaries between adjacent chromatin domains. Insulators bind proteins that modify chromatin and alter regional gene expression. The placement of insulators in the vectors described herein offer various potential benefits including, but not limited to, 1) shielding of the vector from positional effect variegation of expression by flanking chromosomes (*i.e.*, barrier activity, which may decrease position effects and vector silencing); and 2) shielding flanking chromosomes from insertional trans-activation of endogenous gene expression by the vector (enhancer blocking). There are two basic classes of chromatin insulators: (a) barrier insulators that block the encroachment of silencing heterochromatin into adjoining regions of open chromatin that are transcriptionally permissive, and (b) enhancer blocking insulators that prevent enhancer-mediated transcriptional activation of adjoining regions. The sequences that mediate these activities are physically separable and mechanistically distinct (Recillas-Targa *et al.* (2002)). Chromatin insulators do not exhibit inherent transcriptional enhancing or repressing activities on their own. As such, they make ideal elements for reducing the interaction between gene transfer vectors and the target cell genome. Insulators can help to preserve the independent function of genes or transcription units embedded in a genome or genetic context in which their expression may otherwise be influenced by regulatory signals within the genome or genetic context (see, *e.g.*, Burgess-Beusse *et al.* (2002) *Proc. Nat'l Acad. Sci. USA*, 99: 16433; and Zhan *et al.* (2001) *Hum. Genet.*, 109: 471).

[0068] The problems created by insertional mutagenesis of viral vectors are widely known (Nienhuis (2013), Baum *et al.* (2006), Nienhuis *et al.* (2006)) as is the evidence that the risks of genotoxicity can be reduced by the use of chromatin insulators (Arumugam *et al.* (2007), Emery (2011), Evans-Galea *et al.* (2007), Rivella *et al.* (2000), Emery *et al.* (2000), Emery *et al.* (2002), Yannaki *et al.* (2002), Hino *et al.* (2004), Ramezani *et al.* (2003), Ramezani *et al.* (2008)). The presently disclosed subject matter provides novel insulators that are powerful enhancer blocking insulators, and certain insulators additionally possess barrier insulator activity. In vertebrates, the function of enhancer blocking insulators is mediated through the zinc-finger DNA-binding factor CTCF (Gaszner and Felsenfeld (2006), Wallace and Felsenfeld (2007)). In general, these elements are thought to function through physical loop structures, which are established by CTCF-mediated interactions between adjacent insulator elements or through CTCF-mediated tethering of the chromatin fiber to structural elements within the nucleus. The first characterized vertebrate chromatin insulator is located within the chicken β -globin locus control region. This element, which contains a DNase-I hypersensitive site-4 (cHS4), appears to constitute the 5' boundary of the chicken β -globin locus (Prioleau *et al.* (1999) *EMBO J.* 18: 4035-4048). A 1.2-kb region containing the cHS4 element displays classic insulator activities, including the ability to block the interaction of globin gene promoters and enhancers in cell lines (Chung *et al.* (1993) *Cell*, 74: 505-514), and the ability to protect expression cassettes in *Drosophila* (*Id.*), transformed cell lines (Pikaart *et al.* (1998) *Genes Dev.* 12: 2852-2862), and transgenic mammals (Wang *et al.* (1997) *Nat. Biotechnol.*, 15: 239-243; Taboit-Dameron *et al.* (1999) *Transgenic Res.*, 8: 223-235) from position effects. Much of this activity is contained in a 250-bp region. Within this stretch is a 49-bp cHS4 element (Chung *et al.* (1997) *Proc. Natl. Acad. Sci., USA*, 94: 575-580) that interacts with the zinc finger DNA binding protein CTCF implicated in enhancer-blocking assays (Bell *et al.* (1999) *Cell*, 98: 387-396).

[0069] Insulators, such as cHS4, can block the interaction between enhancers and promoters when placed between these elements (Evans-Galea *et al.* (2007), Chung *et al.* (1997), Bell *et al.* (1999), Ryu *et al.* (2007), Ryu *et al.* (2008)). Several studies have demonstrated the ability of the cHS4 insulator to reduce position-effect silencing of gammaretroviral vectors (Evans-Galea *et al.* (2007), Rivella *et al.* (2000), Emery *et al.* (2000), Emery *et al.* (2002), Yannaki *et al.* (2002), Hino *et al.* (2004), Ramezani *et al.* (2006), Yao *et al.* (2003), Nishino *et al.* (2006), Aker *et al.* (2007), Li and Emery (2008)), and lentiviral vectors (Bank *et al.* (2005), Arumugam *et al.* (2007), Puthenveetil *et al.* (2004), Evans-Galea *et al.* (2007), Ramezani *et al.* (2003), Aker *et al.* (2007), Ma *et al.* (2003), Chang *et al.* (2005), Pluta *et al.* (2005)). Those appropriately designed studies demonstrated that inclusion of the 1.2 kb version of the cHS4 insulator increased the likelihood and/or consistency of vector transgene expression in at least some settings (Arumugam *et al.* (2007), Emery (2011), Evans-Galea *et al.* (2007), Emery *et al.* (2002), Yannaki *et al.* (2002), Hino *et al.* (2004), Ramezani *et al.* (2006), Aker *et al.* (2007), Li and Emery (2008), Pluta *et al.* (2005), Jakobsson *et al.* (2004)). Nevertheless, the degree of protection afforded by the cHS4 insulator is far from complete. In addition, the inclusion of the 1.2 Kb cHS4 can adversely

affect vector titers while the smallest cHS4 core has been proven ineffective (Aker *et al.* (2007), Jakobsson *et al.* (2004)). By contrast, the insulators of the presently disclosed subject matter do not affect adversely the titers of viral vectors, and are more powerful and effective than the cHS4 insulator.

[0070] The presently disclosed insulators are identified through genomic approaches, e.g., using genomic approaches to identify insulators that are powerful enhancer blockers as well as barrier insulators of the human genome. The presently disclosed insulators enhance the safety of gene therapy (e.g., stem cell gene therapy, globin gene therapy). For gene therapy of the hemoglobinopathies, powerful enhancers are required to achieve therapeutic levels of globin gene expression. Powerful insulators therefore represent one means to protect the genomic environment from the powerful enhancers of the integrating vectors.

[0071] The presently disclosed insulators possess powerful enhancer blocking activity. For example, and not by way of limitation, an insulator of the present disclosure can reduce the activity of an enhancer element by at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98% or at least about 99%. In certain embodiments, the insulators possess barrier activity in addition to enhancer blocking activity. The presently disclosed insulators substantially decrease the risks of insertional mutagenesis and genotoxicity associated with viral vectors. Furthermore, when a presently disclosed insulator is incorporated into a vector, the insulator does not adversely effect vector titers of the vector. In certain embodiments, the insulators (e.g., insulator A1) increase the *in vivo* expression of the globin gene or functional portion thereof.

[0072] In certain embodiments, the insulator comprises a Transcriptional repressor CTCF binding site, which has the nucleotide sequence set forth in SEQ ID NO: 18, which is provided below:

CACCAGGTGGCGCT [SEQ ID NO:18].

[0073] In one non-limiting embodiment, the insulator has the nucleotide sequence set forth in SEQ ID NO:1, which is provided below, or a sequence which is at least about 95 percent homologous, or at least about 98 percent identical (homologous), to SEQ ID NO:1.. This insulator having the nucleotide sequence set forth in SEQ ID NO:1 is designated as insulator A1.

```
TCCTTCCTTTCTAAATGACGAGAGAGACAGAAGAATTCTTCAAGGTTAGTGTGTCCAGCATG
CAACCTTTTCCTTCCTGGATGAGCATCCCTGGAGTAGGAGAGCCAGCCTGCCTCCTGCGCTGG
CACAGAGCCCGGTTCCCTAGACAACTGCCTCTCCAAATCTGATGTCCAGCGCCACCTGGTGT
CCACATCAAGCAGACACAATTAATAGTCAACCTGTTTCAGGAAAAGTGTGAGGGGGGAAAAAAA
AGAAAGAGGATTTATGAAGGGAAAAGAAAGTTTAGAGGATATGCCACGATTGGCTAG [SEQ
ID NO:1]
```

[0074] In certain embodiments, the insulator comprises a nucleotide sequence as set forth in SEQ ID NO:24, , or a sequence which is at least about 95 percent identical, or at least about 98 percent identical, to SEQ ID NO: 24.

```
CCAATC GTGGCATATC CTCTAAACTT TCTTTTCCCT TCATAAATCC TCTTTCTTTT
TTTTCCCCCT CACAGTTTTC CTGAACAGGT TGAATATTAA TTGTGTCTGC
TTGATGTGGA CACCAGGTGG CGCTGGACAT CAGATTTGGA GAGGCAGTTG
TCTAGGGAAC CGGGCTCTGT GCCAGCGCAG GAGGCAGGCT GGCTCTCCTA
TTCCAGGGAT GTCATCCAG GAAGGAAAGG TTGCATGCTG GACACACTAA
CCTTGAAGAA TTCTTCTGTC TCTCTCGTCA TTTAGAAAGG AAGGA [SEQ ID NO:
24] .
```

[0075] In certain embodiments, the insulator comprises a nucleotide sequence as set forth in SEQ ID NO:25 (which is the reverse complement of SEQ ID NO:1), or a sequence which is at least about 95 percent identical, or at least about 98 percent identical, to SEQ ID NO: 25.

CTAGCCAATCGTGGCATATCCTCTAAACTTTCTTTTCCCTTCATAAATCCTCTTTCTTTTTT
 TTCCCCCTCACAGTTTTCTGAACAGGTTGACTATTAATTGTGTCTGCTTGATGTGGACACC
 5 AGGTGGCGCTGGACATCAGATTTGGAGAGGCAGTTGTCTAGGGAACCGGGCTCTGTGCCAGC
 GCAGGAGGCAGGCTGGCTCTCCTACTCCAGGGATGCTCATCCAGGAAGGAAAGGTTGCATGC
 TGGACACACTAACCTTGAAGAATTCTTCTGTCTCTCTCGTCATTTAGAAAGGAAGGA [SEQ
 10 ID NO:25]

[0076] In certain embodiments, the insulator comprises a nucleotide sequence as set forth in hg18 coordinates 76229933 to 76230115 of chromosome 1.

[0077] In certain embodiments, the insulator comprises a nucleotide sequence between residues 68041 and 68160, or between residues 68041 and 68210, or between residues 68041 and 68280, or between residues 68005 and 68305, of *Homo sapiens* chromosome 1 clone RP11-550H2, GenBank Accession No. AC092813.2, or a sequence at least 95 or 98 percent identical thereto.

III. Expression Cassettes

[0078] The presently disclosed subject matter provides expression cassettes comprising one or more the above-disclosed insulators (e.g., insulator A1). In certain embodiments, an expression cassette comprises at least one insulator having the nucleotide sequence set forth in SEQ ID NO:1, and a globin gene or a functional portion thereof operably linked to a β -globin LCR region.

β -globin LCR region

[0079] The human β -globin gene cluster consists of five genes embedded within one of many olfactory receptor gene arrays (Bulger et al., PNAS (1999);96:5129-5134). The cluster spans over 80kb on chromosome 11p15.4, and includes the five expressed β -like genes and *cis*-acting regulatory elements that direct their stage-specific expression during ontogeny (Forget (2001), Molecular Mechanism of Beta Thalassemia. Steinberg MH et al., Eds. Disorders of Hemoglobin. Genetics, Pathophysiology and Clinical Management, Cambridge University Press, Cambridge). The genes are arranged in the order of their developmental expression (Stamatoyannopoulos et al., (2001) Hemoglobin Switching. In: Stamatoyannopoulos G, et al., Eds. Molecular Basis of Blood Disorders, W.B. Saunders, Philadelphia, PA), 5'- ϵ - γ - $A\gamma$ - ψ η - δ - β -3'. The α -like globin gene cluster (5'- ξ 2- ψ ξ 1- ψ α 2- ψ α 1- α 2- α 1- θ -3') is located very close to the telomere of the short arm of chromosome 16 and spans about 40kb. The expression of genes encoded within these two independent clusters is limited to erythroid cells and balanced so that the output of the β -globin-like chains matches that of the α -chains. This fine tuned balance is regulated at the transcriptional, posttranscriptional and posttranslational levels.

[0080] Developmental stage-specific expression is controlled by a number of proximal or distal *cis*-acting elements and the transcriptional factors that bind to them. In the case of the β -globin gene (*HBB*), the proximal regulatory elements comprise the β -globin promoter and two downstream enhancers, one located in the second intron of β -globin and the other approximately 800bp downstream of the gene (Antoniou et al., EMBO J. (1988);7:377-384; Trudel et al., Genes Dev. (1987);1:954-961; Trudel et al., Mol. Cell. Biol. (1987);7:4024-4029). The most prominent distal regulatory element is the β -globin LCR, located 50-60kb upstream of the *HBB* and composed of several sub-regions with heightened sensitivity to DNaseI in erythroid cells (Forget (2001); Grosveld et al., Cell (1987);51:975-985; Talbot et al., Nature (1989);338:352). The most prominent property of the LCR is its strong, transcription-enhancing activity. An exemplary nucleotide sequence of the human β -globin region on chromosome 11 is set forth in SEQ ID NO:19 (GenBank Access No.: NG_000007.3), which is provided below:

ggatcctcacatgagttcagtatataattgtaacagaataaaaaatcaattatgtattcaagttgctagtgcttta
 agaggttcacatttttatcttaactgattatcacaaaaatacttcgagttacttttcattataattcctgactacac
 atgaagagactgacacgtaggtgccttacttaggtaggttaagtaatttatccaaaaccacacaatgtagaaccta
 5 agctgattcggccatagaaacacaatatgtggtataaatgagacagagggatttctctccttcctatgctgtcaga
 tgaatactgagatagaatatatttagttcatctatcacacattaaacgggactttacatttctgtctgttgaagattt
 ggggtgtggggataactcaaggtatcatatccaagggtatggaagcaggtgactctaacagaaagggaaggat
 gttggcaaggctatgttcatgaaagtatatgtaaaatccacattaagcttctttctgcatgcattggcaatgttta
 tgaataatgtgtatgtaaaagtgtgctgtatatattcaaaagtgtttcatgtgcctaggggtgtcaaatactttgagt
 10 ttgtaagtataacttctctgtaatgtgtctgtaatatctctatttacttgatttctcaataagtaggtatcatagtg
 aacatctgacaaatgtttgaggaaatatttagtgtttacctattccacaaaatttattaaatgcctaactctgtatc
 agatatacaattatctggcgaaatctgtaattcctaatttaaacagctgtgtagcctaatttagggataaaggcatg
 caaacccataatttgtgtaggttgaaatgagctatagaaaaatgcagtataatttatcagaagtcctttagggtcatg
 aaaaggaatggtaactgacactgccagggactcatatgtaagagataactaatgtgaagtgactttaaggagaa
 attagcagaagttttctttccatgtctcctcatcatgttacaataacggaagagattaaacaacaaatacattta
 15 gacagcaatgtttatcctgggttagatgttttaatacctaatactatcttgagtggttaaaatgcatttgctcacctac
 tttaaaatataaatgaaggttaggaacctgtagatcacaaaaagttggagaaaaaaagacaataaagatgacaaaaat
 ctattaatccttgatagaaaaatgagaagagataaaaacactggtttacataaagaaaaataagatggatagatagcag
 atccttataaaaagtgtataatttgagaaaaaaaataactccataattctgagtttcttcacataaaaataatacaaatct
 gctgtggtaagttacaaagagatagattttttatcattatataaaaagatatttttaaacagagttataacaacaaagg
 aacagactatgtcatataattctcacttatcactataaacactcagaaaaatctgcaaaatcatttcatagcattt
 20 taaatagttaggaataatgtagaaaaactgaaacagttctaagtttccacaaacttagagtgctcaaatgttgccatt
 acctaactttacctgcaaatattttatacaaaatttgacatgtctactctagtcaaaaaatataatgtacattatgggta
 ttttctgtgtgtaacttgggtctagttgtcttcttcagaaatagcctctatttttgatttacctgataaaaatcaca
 ttctctccaaagccttctaaataacttcagactaaactacttttttagtacatctaagaagaaaagagttttgtctc
 ttatccacctctgagtcacaaaagcagcatgtccatcaattgggtacatagttcccacagccccacttagctctggat
 25 tggagtttctacttggcattgtttgcaactacatggacgtaaaatgcattggatttctcttgaaaaatgtttctgcc
 tgatgttctctgaaagagactaaccttccctcgcttgcagagaaagactcgtgtaatccttgacaatgtcatctc
 atctattttatcccatgtctaccatgtgtgaccttcatgtcttctgcttaagccccacatcctcaatctacaca
 ctaggatagtataaaagtaatagtaataatagtagtaatagtaataacaatacaatgattatggcttatactatac
 acaagacactgttgatataattatttcattttagtattcacagtaactctgtgcctcaagtactattgtaataacctt
 30 taagaggaggaactgaggcacagggccctaaagtaatatccaaagatgaagtggctactaactgacagagggcat
 aattcaactcatgatatttggctctagaatacatgtctctgaatcattatacaataataattcatgaggaaacattt
 tttaaagcctaagttatttgcctctgaaataagacataaatttgggggtgagaaagcttagattccatgaagtattaca
 gcatttggtagtcttttgcactccaggtcttattttactgcttaaacataataaaaacatatggttcagtatgcc
 tttgattttacaataatattcctgttatttttgggaagcacaggggtgtgggataatgctaattactagtgattagta
 35 ttgagaggtgacagcgtgctggcagtcctcacagccctcgctcgctcttgggcgctcctctgcctgggctcccaca
 ttgggtggcacttgaggagcccttcagccggcgcgtgcactgtgggagcccttttctgggctggccaaggccagagc
 cggtcctcctcagcttggcaggaggtgtggagggaacagacgcgggcaggaaacgggctgtgcgcgctgtctgaggga
 gttccgggtgggcatgggctccgaggaccccgactcggagccgcagccggccccaccggccgaggcagtgagg
 ggcttagcacctgggcccagcagctgtgtgtcaattcctcgccggccttagctgccttctgcccgggaggggct
 cgggacctgcagcgcgcctatgctgagcctccccaccttcatgggctcctgtgcggcccgagcctcgccgacgagc
 40 gccgccccctgctccagggcacccagtcctcatcgaccacccaagggtgaagagtgccgggcgcagggcaggggact
 ggcaggcagctccccctgcagcccagggtgcccgatccactgggtgaagccggctaggtcctgagtttgcctgggga
 tgccaagaaccttatgtctagataagggttgtaaatcaccaattggcactctgtatctagctcaaggtttgta
 aacacaccaatcagcacctgtgtctagctcagggtttgtgaatgcaccaatcaacactctatctagctactctgg
 tggggccttgagaacctttatgtctagctcagggtttgtaaatcaccaatcggcagctctgtatctagctcaagg
 45 tttgtaaacacaccaatcagcacctgtgtctagctcagggtttgtgaatgcaccaatcaacactctgtatctagc
 tactctgggtggggacgtggagaacctttatgtctagctcagggtttgtaaatcaccaactcggcagctctgtatcta
 gctcaaggtttgtaaacacaccaatcagcacctgtgtctagctcagggtttgtgaatgcaccaatcaacactctg
 tatctagctactctgggtggggacttgaggaaacctttgtgtggacactctgtatctagctaatctgggtggggacgtg
 gagaacctttgtgtctagctcatggattgtaaatgcaccaatcagtgccctgtcaaaacagaccactgggctctac
 caatcagcaggatgtgggtggggccagataagagaataaaagcaggtgcccagagccagcagtggaacccgctcg
 50 ggtccccttcacactgtggaagcttgggtcttctcgctcttgcataaaatcttgctgctgctcactgtttgggtc
 tacactgcctttatgagctgtaacgctcacccgcgaaggtctgcagcttcaactcttgaagccagcagaccacgaac
 ccaccgggagggaacgaacaactccagaggcgccgcttaagagctggaacgttcaactgtgaaggtctgcagcttca

55

ctctcgagccagcgagaccacgaacccatcagaaggaagaaactccgaacacatccaaacatcagaacgaacaaac
tccacacacgcgagcctttaagaactgtaacactcaccacgaggggtccccggttcattcttgaagtcagtgaacc
aagaacccaccaattccggacacagtatgtcagaacaatatgagtcactaaatcaatatacttctcaacaatttc
caacagcccttgcaattaaacttggccatgtgactggttgtgactaaaataatgtggagataataatgtgttactcc
5 ctaaggcagagtgcccttctatcatctcttcttcccttctctatgtggcgagaaagttaaagattctgaaatgataa
agtcaatcacaggaaggcaccttggactcctggcccactgcttggaggagagcactcaggaccatgaacatctgact
gtgacgtagcaataaagaacccacggtttcatatgaaactgcttaaaattatggcagacagtcagtctgtttttagt
tgcacatttgtctttatttgtgcttgttttgcctcacacatcaactcaccagggcctacattctctgataatgca
atltcaagttctttacaggccgagaaaaatgaatctgaatctctgacctccaaaagtgatcaagatatttttagtt
caggtccaaaaattttctcattttcataggttttctctgattgatcattattcatgatttgcaaggaatcattcaa
10 tgttttctaaatctattactgcatcctgacacatatgacattttaactatgttccagatttttgatgaagagt
aaattttaaatgttttcaccacaaaaataagtatgtgaagtgtgtgatttgttaattagccttatttaaccattt
aatattgtacacgtacaccaaagcatcatgttgtaccccatgaatacacacaattattatttgtcaattttaaatg
aataataaaaaatacaaaagcattagcctctgcatgtgctttacccggtcacctcagcgtgactaacgcaaaaa
acgttctatttctccttcaaaaactccctattctgtagcctcttctgtagattctctatccctcctgtttct
15 ctacgttatttctatgtggtatcatcaccatcctggacaacatcaggacagatatccctcaccaagccaatgttct
ctctatgttggctcaaatgtccttgaactttcctttcaccacccctttccacagtcacaaaggatattgtagttaa
gctcagagttcagcttttaagcttctgacaaattattcttctctttaggttctcctttatggaatcttctgtac
tgatggccatgtcctttaactactatgtagatatctgctactacctgtattatgcctctacctttatttagcagagt
tatctgtactgttggcatgacaatcatttgttaatatgacttgcccttctcttttctgctattcttggatcaaatgg
ctcctctttctgtcctctcattttctcctgccttcaacttggacgtgcttcacgtagtctgtgcttatgactggat
20 taaaaattgatatggacttactcctaattgtgttctgataaatatgggtttttagtttggccattatttattctctatgca
ttgatctggagaaggtctcaatccttttactccttgggaaaaatatctgtaaaccttctgttactcctctatag
caatttcagttttaggtctagtaagcatgaggatgcctccttctctgatttttccacagctctgttgggtcacagaata
acctgagtgattactgatgaaagagtgagaatgttattgtagtcacaatgacaaaaaacaaacaactacagtcaa
aatgtttctctttttattagtggattatatttcttgacctatctctggcaggactctttagagaggttagctgaagc
25 tgcgtttatgaccactagagggaagaagatacctgtggagctaatgggtccaagatggttgagcccaagcaaggaa
gttgttaaggagcccttttgattgaaggtgggtgccccaccttacagggacaggacatctggatactcctcccag
tttctccagtttccctttttcttaatatatctcctgataaaatgtctatactcacttccccatttctaataataaa
gcaaaggctagttagtaagacatcaccttgcattttgaaaattgcatagactttcaaaattatttcacatcctggt
ccttctttatttcaagagtcagaaaatggcaacattcactttagtccaatgtaatggaaagagctctttcaagaga
cagagaaaaagaataatttaatttcttcccaacactccttccctgtctcttaccctatcttcttcttcttacc
30 tccccatttctctctctcatttctcagaagtatattttgaaaggattcatagcagacagctaaaggctgggtttttc
taagtgaagaagtgatattgagaaggtagggttgcattgagccctttcagtttttttagtttatatacatctgtattg
ttagaatgttttataatataaataaaattatttctcagtttatatactagctatgtaacctgtggatatttccctaa
gtattacaagctatacttaactcacttggaaaactcaataaataacctgcttcatagtttattaataaggattaa
gagataatgcccataagattcttatataaacagataaatacacacacacacacattgaaaggattccttact
35 ttgtgttaggaactataataagttcatttgatgcatattatcattaaacttcaatttcaacactagaaggcaggtat
tatctaaatttctacttgatcatcctccaaactcataagataattaaattgcctcttctgtcatatttattcaaaa
gggtaaactcaactatggcttgtctaattttatatatcacctactgaacatgacctattgtgatattttataaa
aattattctcaagttattatgaggatgttgaaagacagagaggatgggggtgctatgccccaaatcagcctcacaat
taagctaagcagctaagagtccttgagggttagtgtagggaccacagggttaagggggcagtagaattatactcca
cttagtttcatttcaacaatccatacacacacagccctgagcacttacaattatactacgctctatactttt
40 gtttaaatgtataaataagtggtgaaagaatagatagatagatagacagatagatgatagatagaataaatgctt
gccttcatagctgtctccctaccttgttcaaaatgttctgtccagaccaaagtaccttgccttcaacttaagta
caattcctaggttatattctgattgcaagggaagtcaaaagatgtgaaaaacaatttctgaccacaaactcagct
ttgtagatgactagatcaaaaaatttcagccatatcttaacagtgagtggaacaggaaactcctcttctccctaca
tctgagatcccagcttctaagaccttcaattctcactcttgatgcaacagaccttggaagcatacaggagagctga
acttgggtcaacaaaggagaaaaagttgttggcctccaaaggcacagctcaaacctttcaagccttctctaattcta
45 aaggtaaacaagggtctcatttcttggagaacttcaggggaaaatagacaaggacttgcctgggtgcttttggtaggg
gagcttgaccttcccccttcttgaggaaatatttatccccaggtagttcccttttgcaccagtggttctttga
agagacttccacctgggaacagttaaacagcaactacaggggccttgaactgcacacttccagtcgggtcctcacag
ttgaaaagacctaaagcttgtgctgtattaaagccttttgggtcataaaacattgaaattctaatctcctctcaacc
ctacagtcacccatttggatattaaagatgtgtgtctactgtctagtatccctcaagtatgtcaggaattagct
cattttaaatgagtctgcaagccaggagtggtgctcatgtctgtaattccagcacttgagaggtagaagtgaggagga
50 ctgcttgagctcaagagtttgatattatcctggacaacatagcaagacctcgtctctacttaaaaaaaaaaaaaa
attagccaggcatgtgatgtacacctgtagtccagctactcaggaggccgaaatgggaggatcccttgagctcag
gaggtcaaggctgcagtgagacatgatcttgcactgcactccagcctggacagcagagtgaaaccttgcctcacg
aaacagaatacaaaaaaaacaaacaaaaaactgctccgcaatgcgcttcccttgatgctctaccacataggtctgg
gtactttgtacacattatctcattgtcttataatgtttagatttaattttgtaatttgatatttattcttagaaa
55 ctgaggcctcaagatgataacttttatttctggacttgaatagctttctctgtattcaccatgttgttaactt
tcttagagtagtaacaatataaagttatttggagttttggcaaacacagcaaacacacagcccatatagacattg
atgtgaaattgtctattgtcaatttatgggaaaaaagtatgtacttttctactaagccattgaaacagggaataa

cagaacaagattgaaagaatacattttccgaaattacttgagtattatacaaaagacacagcagctggacctgggagg
 aggggtattgtccatgactggtgtgtggagacaaatgcagggtttataatagatgggatggcatctagcgcaatgac
 tttgccatcacttttagagagctcttggggacccagtagacaaagaggggacgcagggtatagttagacatctcat
 tctttttcttagtgtgagaataagaatagccatgacctgagtttatagacaatgagcccttttctctctccactc
 5 agcagctatgagatggcttgcctgctctctactaggtgactcactccaaggccagcaatggcgagggtctgt
 tcagggtcttgatagcactatctgcagagccagggtcgagaagggttgactccagagactctccctccattccc
 gagcagggtttgcttatttatgcatttaaatgatataatttttaaaagaataaacaggagactgccagccctg
 gctgtgacatggaaactatgtagaataattttgggttccatttttttcccttcttccagtttagaggaaaagggtc
 cactgcacatacactagacagaaagtcaggagctttgaaatccaagcctgatcatttccatgtcatactgagaaagt
 10 ccccccccttctctgagcctcagtttctctttttataagtaggagctctggagtaaatgatttccaatggctctcat
 ttcaatacaaaaatttccgtttattaaatgcatgagcttctgttactccaagactgagaaggaaatgaaacctgaga
 ctcatgactggcaagatgtccccagaggctctcattcagcaataaaatttccaccttcccccagggccactgagt
 gtcagatttgcactgactagtccactgtgtgtaaaaaggaggtgcttcttcttctgtattctcacataccttttag
 gaaagaacttagcacccttccacacagccatcccaataactcatttcagtgactcaacccttgactttataaaag
 tcttgggcagtatagagcagagattaagagtacagatgctggagccagaccacctgagtgatttagtgactcagttt
 15 ctcttagtagttgtatgactcagtttcttcatctgtaaaaatggagggttttttaattagtttgttttgagaaagg
 gtctcactctgtcaccctaaatgggagtgtagtggcaaaatctcggtcactgcaacttgacttccagggtcaag
 cggctctccacctcaacatcctgagtagctggaaccacagggtacacaccaccatacctcgctaattttttgtatt
 tttgtagagatgggggtttcacatgttacacaggatgggtctcagactccggagctcaagcaatctgccacctcag
 ccttccaaagtgtgggtattataagcatgattacaggagttttaacaggctcataagattgtttcgcagcccgagt
 20 gaggttatacatgcaaaaggtttaaagcagtgacttataaatgctaactactctagaaatgtttgctagtattttt
 tgtttaactgcaatcattcttctgtgcaggtgaaaactagtgttctgtactttatgccattcatctttaactgtaa
 taataaaaaataactgacatttatgaaaggctatcagagactgtaattagtgctttgcataaattaatcatattta
 actcttggtattctttcaggtagatactattattatccccattttactacagttaaaaaaaactacctctcaacttgc
 tcaagcatacactctcacacacacaaacataaactactagcaaatagtagaattgagatttggtcctaattatgtc
 tttgtcactatccaataaataatttattgacatgtacttcttggcagctctgtatgctggatgctgggatacaaa
 25 atgtttaaatttaagctccagttctgtcttccaaaggcctccaggcccaagttatccattcagaaagcatttttta
 ctctttgcatttccactgttttttctaagtgactaaaaattacactttattcgtctgtgtcctgtctgggatgat
 agtctgacttttccataacctgagcctaaccatccctgacatcaggaagactacaccatgtggagaagggtggtggt
 tttgattgctgtctgtcttccagtttagatggttaactttgtgaaagtgaaaaactgtggctctctggttgactgttaga
 gttctggcacttgtcactatgcttatttttaacaaatgcatgaatgcttcagaatatgggaatattatcttctgg
 aatagggaaatcaagttatattatgtaacccaggattagaagattcttctgtgtgtaagaatttcataaacattaag
 30 ctgtctagcaaaaagcaagggttggaataatctgtgagctcctcaccatatagaaagcttttaacccatcattgaa
 aatccctataggggattttctacctgagcaaaaaggctggctcttgattaatttccaaaactcatatagctctgagaa
 agtctatgctgttaacgttttcttgtctgtaccccatcatatgcacaacaataaatgcaggcctaggeatgactg
 aaggctctctcatabaattcttgggtgcatgaatcagattatcaacagaaatgttgagacaaactatggggaagcagg
 gtatgaaagagctctgaatgaaatggaacccgcaatgcttctgcccattcagggtccagcatgtagaatcttgg
 35 ggctttgtgaagactggcttaaaatcagaagccccattggataagagtagggaagaacctagagcctacgctgagc
 aggtttccttcatgtgacagggagcctcctgccccgaacttccagggtatcctctcttaagtgtttcctgctggaat
 ctctcacttctatctggaatgggtttctccacagtcacagccctggctagttagaaagagttacccatgcagaggc
 cctcctagcatccagagactagtgttagattcctactttcagcgttggaacacctggatccacttgcccagtggt
 cttccttagtttctaccttgcaccttgatcctcctttatcttctgaacctgtgagatgatctatgtggggaga
 atggcttctttgagaacatcttcttctgtagtggtgcccctcattcccactttaatatccagaatcactataa
 40 gaagaatataataagaggaataactcttattataggttaagggaataaagaggcatacgtgatggatgagtaa
 agagagagggaaggatgaatggacgataaaatctactactatttggtagagacctttatagtctaatcaattttg
 ctattgttttccatcctcacgctaactccataaaaaaacactattattatcttatttttgccatgacaagactgag
 ctcaagaagagtcaggcattttgcctaaggtcggacatgtcagaggcagtgccagacctatgtgagactctgcagcta
 ctgctcatgggcccctgtgctgcaactgatgaggaggatcagatggatggggcaatgaagcaaaaggaatcattctgtg
 45 gataaaggagacagccatgaagaagtctatgactgtaaatttgggagcaggagctcttaaggacttggaattcaag
 gaattttgactcagcaaacacaagacctcaggtgacttttgcagagctgggtgtgcccagatgtgtctatcagagggt
 ccagggaggggtgggtgggtcagggctggccaccagctatcagggcccagatgggttataggctggcaggctcag
 ataggtgggttaggtcaggttgggtgggtgctgggtgaggtcactgactcccaggagccaggagagatagaccatgag
 50 agagggcagacatgggaaagggtgggggagggcacagcatagcagcatttttcttactactacatgggactgctc
 cctataccccagctaggggcaagtgccttgactcctatgttttcaggatcatcatctataaagtaagagtaata
 attgtgtctatctcatagggttattatgaggatcaaaaggagatgcacactctctggaccagtggtcctaacagttca
 ggacagagctatgggttctctatgtatgggtcagtggtctcaatgtagcaggcaagttccagaagatagcatcaac
 cactgttagagatatactgccagctctcagagcctgatgttaatttagcaatgggctgggacccctcctccagtagaa
 ccttctaaccagctgctgcagtcaggatgaatgcagctgggttagacttttttaaatgaaagcttagctttcatta
 aagattaagctcctaagcagggcacagatgaaattgtctaacagcaactttgcatctaaaaaaatctgacttcac
 55 tggaaacatggaagcccaaggttctgaacatgagaatttttaggaatctgcacaggagttgagaggggaacaaga
 tgggtgaaggagactagaaccacatgagagacacaggaataatgttagatttaggctggagggttaaatgaaagaa
 gtgggaattaataacttactgaaatcttctatgtcaggtgcccattttatgatatttaataatctcattacatat
 ggtaattctgtgagatagtattattgaacatactataatttaataactaatgataagtaaacacctcttgagtactta

gtatatgctagaatcaaatttaagtttatcatatgagggcgggcacgggtggctcatatatgggattacatgcctgt
 aatcccagcactttgggagggccaaggcaattggatcacctgaggtcaggagttccagaccagcctggccaacatgg
 tgaaccccttctctactaaaaaatacaaaaaatcagccagggtgtgggtggcacggctctataatcccagctactca
 5 ggaggtgagggcaggagaatcacttgaaccaggaggtggaggtgcagtgagctaagattgcaccactgcactcc
 agcctaggcgacagagtgagactccatctcaaaaaaagaggtttattatatgaattaaacttagttttact
 cacaccaatactcagaagtagattattacctcattttatgtatgaggagcccaatgtacttgtatgtatgaact
 tattgaaagcacaaagtaataagtagacaattagtaattagaagtcagatggctgtgagctctcctactgtctacat
 tacatgagctcttattaaactggggactcgaaaatcaaagacatgaaataatttgtccaagcttacagaaccacca
 gtagtaaggctaggtatgtagaccagttctgtactctgaagacagtggtttttccacagcaaaacacaaactca
 10 gatattgtggatgcgagaaattagaagtagatattcctgcctctgtggcccttgcttcttacttttacttcttctg
 attggaagttgtgttccaagccacagttgcagaccatacttctcaccataattgcatttcttcaggaaagtttg
 agggagaaaaaggtaaagaaaaatttagaacaacttcagaataaagagattttctctgggttacagagattgtc
 atatgacaaattataagcagacacttgagaaaactgaaggcccatgcctgcccataattaccccttgacccttggt
 caagctgcaactttgggttaaaggagtggtttatgtgttatagtgttcatttactcttctgggtctaaccattggct
 ccgtcttctcctgcagtgacctcagtgccctcagaacatacatatgtttgtctagttaaagtttgtgtgaaattc
 15 taactagcgtcaagaactgagggccctaaactatgctaggaatagtgtgtgtgtgtgtgataggtacacaagaaa
 tgagaagaaactgcagattctctgcactccctttgcccgggtctgacaacaaagtttcccaaaattttaccaatgc
 aagccatttctccatatgtcaactactttaaaaatcatttgggggttcacattgtcttctcactctgtaaaaaagaat
 ggaagaactcattcctacagaactccctatgtcttccctgatgggctagagttcctcttctcaaaaattagccat
 tattgtatttctctotaagccaaagctcagaggtcttgtattgcccagtgacatgcacactgggtcaaaagtaggct
 aagtagaagggtactttcacaggaacagagagcaaaagaggtgggtgaatgagaggttaagtgaagaaagacaaat
 20 gagaagttacaacatgatggtctgtgtctaaatatctcctaggggaattattgtgagaggtctgaatagtgtgtga
 aaataagctgaatctgtgtccaacattaacagtcgaagaaatacctccgaataactgtacctccaattattctttaa
 ggtagcatgcaactgtaatagttgcatgtatatatttatcataactgtaacagaaaaacttactgaatatata
 ctgtgtccctagtctctttacacaataaactaatctcatcctcataattctattagctaatacatattatcatccta
 tatttccagagacttcaagaagttaagcaacttgctcaagatcatctaagaagtaggtggtatttctggggtcattt
 ggccctcctaatctctcatggcaacatgggtgcttaaagtggtgattgaccttaattcatcagggatggggtcata
 25 ctactgcagaccttaactggcatcctcttcttcttattgtgatctgacctgacctagtagactttagaaatttctga
 tgagaaaggagagaggaaggaagggcagagctgactgtgatgtgatgaaggtgaccttctcatctgggtaccagtgg
 ggctctaaagactaaagtcactctgtctcactgtgtcttagccagttccttacagcttgccctgatgggagatagag
 aatgggtatcctccaacaaaaaataaattttcatttctcaaggtccaacttatgttttcttaatttttaaaaaaa
 tcttgaccatttctccactctctaaaataatccacagtgagagaaacattcttttccccatcccataaatacctct
 30 attaaatatggaaaatctgggcatggtgtctcacactgtaatcccagcactttgggaggtgaggtgggtggact
 gcttggagctcaggagttcaagaccatcttggaacaacatgggtgataccctgacctacaaaaagtaaaaaattag
 cctggcatggtggtgtgcacctgtaatcccagctattaggggtggctgagggcaggagaattgcttgaaccgggagg
 cggaggttgagtgagctgagatcgtgccactgcactccagcctgggggacagagcacattataaactggttat
 ttttacttgactcttgtgggaataagatacagtgttttattcttatttatgattcaagcactgaaaatagtggt
 tagcatccagcaggtgcttcaaaaccatttgcgtgaatgattactatactttttacaagctcagctccctctatccc
 35 ttccagcatcctcatctctgattaaataagcttcagtttttctttagttcctgttacatttctgtgtgtctccatt
 agtgacctcccatagtcacaagcatgagcagttctggccaggccctgtcggggtcagtgccccaccccgccctct
 ggttctgtgtaaccttctaagcaaaccttctggctcaagcacagcaatgctgagtcagtgatgagtcagtgaggc
 ttagggtgtgtgcccagatgttctcagcctagagtgatgactcctatctgggtccccagcaggatgcttacagggc
 agatggcaaaaaaaggagaagctgaccacctgactaaaaactccacctcaaacggcatcataaagaaaatggatgc
 40 ctgagacagaatgtgacatatctcagatatattatttctgaaatatatatatatatacacatatacgtatata
 tatatatatatatatatttgtgttatcaattgccatagaatgattagttattgtgaatcaaatattttatcttgc
 ggtggcctctatacctagaagcggcagaatcaggcttttattaatacatgtgtatagatttttaggatctatacaca
 tgtattaatatgaacaaggatatggaagaggaaggcatgaaaacaggaaaagaaaacaaaccttgtttgccattt
 taaggcaccctggacagctaggtggcaaaaggcctgtgtgttagaggacacatgctcacaacggggtcagatc
 tgacttgggggtgctactgggaagctctcatcttaaggatacatctcaggccagttctgtgtcattaggaagatgta
 45 ggcaactctgatcctgagaggaaagaaacattcctccaggagagctaaaagggttcacctgtgtgggttaactgtga
 aggactacaagaggatgaaaaacaatgcagacagacataattgcttgtgggagaaaaaacaggaggtcaaggggat
 agagaaggcttccagaagaatggcttgaagctggctctctgttaggagttcacagtgggcaagatgttccagaatg
 tgacatgacttaagggaactatacaaaaaaggaaacaaatttaaggagaggcagataaaattagttcaacagacatgcaa
 ggaattttcagatgaatgttatgtctccactgagcttcttgagggttagcagctgtgaggggttttgaggccaggga
 cccattacaggacctcacgtatacttgacactgtttttgtattcatttgtgaatgaatgacctcttgtcagctcta
 50 ctcggtttcgctgtgaatgaatgatgtcttgtcagcctacttgggttctgctaagagcacagagagaagatttagtg
 atgctatgtaaaaaacttcccttttgggtcaagtgatgtttgtgatagaaatgaagacaggctacatgatgcata
 ctaacataaacacaaaacattaaagaaaggaaatcaacctgaagagtattttatacagataacaaaatcacagagtgta
 gttaaatgtgtaataactgtggcacaggtggaatatgagccatttaaatcacaaattaaatagaaaaaaaacagt
 ggggaaaaaattccatggatgggtctagaagactaagacttattttaggttgagtgggcaggtgtttaaagggtgata
 55 tcagactaaacttgaaatatgtggctaaataactagaataactcttatttttctgtatcagaatagcagatatag
 cttgatggccccatgcttgggttaacatccttgcgtgttctgacatgaaatccttaatttttgacaaaggggtat
 tcattttcattttatatattgggctagaataattatgtagatggctcctgaggaaaagtttatagcttgcctatttctct

5

10

15

20

25

30

35

40

45

50

55

aggcagccacagaatttatgacttgtgatatccaaagtcattcctggataatgcaaaatctaacacaaaatctagta
gaatcatttgccttacatctatttttgttctgagaatatagatttagatacataatggaagcagaataattttaa
ctggctaatttagaatcctaagcagctcttttctatcagtggtttacaagcctgtttatattttctatttta
5 aaaaataaaaataaagtaagttatttgtggtaaagaatattcattaaagttatttttcttagataaataccatgaaa
aacattcagtgaaagtgaagggcctactttacttaacaagaatctaatttatataattttctactaataagcatct
aagaacagttacaatatttgaactcttcagggttaacatatgtcataaattagccagaaagatttaagaaaaattgg
atgtttcttgttttaattaggtactcttcaggttttagaattctgcctagacacttaagaaattacaattgctaaa
gcaaaccccaacaggcaggaattaatcttcatcgaattttgggtgtttctttctaaaagtcctttacttaactaa
10 cttagacatacatagatttttttactaatttttaattatatagacaataaatgaatatcttactgattacttt
ttctgactgtcctaattctttctgactatctcctggatggccataacacttatctctctgaactttgggcttttaata
aggaaagaaaaagcaataatccatttttcatggtatctcatatgataacaaaataaaatgcttaaaaatgagcagg
gaagcaatttatcttgaaaccaacagcatcgaagcaataatgagactgccgcagcctacctgacttctgagtcag
gatttataagccttgttactgagacacaaacctgggcctttcaatgctataacctttcttgaagctcctccctacc
acctttagccataaaggaacatggaatgggtcagatccctggatgcaagccaggcttggaaccataggcagtaagg
15 agagaagaaaaatgtggctctgcaactggctccgagggcagcaggagaggtcaaccctactctgaaactctgaag
aagactggtgtccatcactctgaattgggaagaatgatgggattcccatagggctgtttttagggaagaaacctgttc
tccaaactcttggccttgagataacctggctcttattccttggactttggcaatgtctgacctcacattcaagttc
tgaggaagggccactgcttccatactgtggatctgtagcaaatccccctgaaacccagagctgtatcttaattg
gttaaaaaaaatttatattatctcaacgactgttctctctgagttagcgaagctcagcttgggtcaagctacaagca
gctgagctgctttttgtctagtcattgttcttttatttcagtggtacaaatcagttctttccaaacctaggatctt
20 gtcttctaggtctatatattttgtcccaggaagtcttaactctggggtccacagaacactagggggctgggtgaagt
tatagaaaaaaaactgtatttttactttacatgtaactgaaatttagcatttttcttactttgaaatgcaaggag
aaactcagaatgacatcactcagtaacctattgcatagtataaagagaacacacagatattttctactacacatag
gtattgcagatctttttgtttttgtttttgtttgagatggagtttctgctcttatggccaggctggagtgcaagt
catgatttcggctcactgcaacctccccctcctgcattcaagcaattctcctgcttggcctcctgagtagctggg
gattacaggcaacctgccaccatgccagtcataatttttgtatttttagtagagatggggtttcgccatgttggccag
25 gctggctcttgaactcctgacctcagatgatctgccgccttggcctcctgaagtgtgggattataggtgtgagcc
accacgcttggccattgcagatatttttaattcacatttatctgcatcactacttggatcttaaggtagctgtag
accaatcctagatctaatgctttcataaagaagcaaatataataaatactataaccacaaatgtaattgtttgatgt
ctgataattgataattcagtgtaattaaacttagcactcctatgtatatttttagtgcaataaaaaacataattttt
tagcacttgcagcttgcacaaactggcctgtgcacaaaaaaagttaggaattcctggttttctgtgttagcca
30 atgggtagaatatatgctcagaagaataccatttggttaatagctaaaagaaaatggagtagaaattcagtggcctg
gaataataacaatttgggcagtcattaaagtcagggtgaagactcttggaatcatgggagaaaaagcaagggagacatt
cttacttggccacaagtgtttttttttttttttttttttatcacaaaacataagaaaaataataaataacaaagtca
ggttatagaagagagaaaaacgctcttagtaaaccttggaatatggaatccccaaaggcacttgacttgggagacagga
gccatactgctaagtgaaaaagacgaagaacctctagggcctgaacatacaggaaatttgtaggaacagaaattcct
agatctgggtggggcaaggggagccataggagaagaaatggtagaaatggatggagacggaggcagaggttgggcag
35 atcatgaggtcaagagatcgagaccatctcggcaaacatgggtgaactcccgctctactaaaaataaaaaaattag
ctgggcagtgctgagctgcgcctgtagtccagctgctcgggagctgaggcaggagaatcggttgaacccgaggag
cgaaaggttgagtgagctgagatagtgccattgcactccagctctggcaacagagtgagactccgtctcaaaaaaa
aaaaaaaagaaagaaagaaaaagaaaaagaaaaaataaatggatgtagaacaagccagaaggaggaactg
ggctggggcaatgagattatgggtgatgtaagggaactttatagaattacaatgctggaatttgtggaactctgct
40 tctattattcccccaatcattactctgtcacattgatagtttaataatttctgtgaatttattccttgattctaa
aatatgaggataatgacaatggattatataagggcagattaaagtatatagcatgagcaatattcttcaggcacatg
gatcgaattgaatacactgtaaatcccaacttccagtttcagctctaccaagtaaaagagctagcaagtcataaaa
tggggcatacacagaaaaaaaaggacactagaggaaataataacctgactcctagctgattaatatcagtt
cactttttctgtttctgttgacaaattctggcctttaataatttttaggatttttaggctcttcagctccctcccc
gtgagaagtataagcaggacagacaggcaagcaagagagagcccaggcaataactcacaaagtagccaatgtccc
ctgtggtcatagagaaatgaaaagagagaggattctctggaagcactggatgtaatttttctgtctgtcctctct
45 agggaatcaccccaaggtactgtactttgggattaaggcttttagtccactgtggactacttgctattctgttcag
tttctagaaggaactatgtacggtttttgtctcctagagaaactaaggtacagaagtttggttacaatgcactc
cttaagagagctagaactgggtgagattctgttttaacagctttattttcttttcttggcctgtttttgtcact
gtcaccacctttaaggcaaatgttaaatgcgctttggctgaaacttttttctctattttgagatttgcctctttat
atgaggtcttcttggaaaagggaatgggagagatggatatacttttgggaagtatgaagagggttaaaaaagggg
50 acaaatggaatttggttgtgagatgagtagaggagcaacaaaaagagcctcaggatccagcacacattatcaca
aacttagtgtccatccatcactgctgacctctccggacctgactccaccctgagggacacaggtcagccttgac
caatgacttttaagtaccatggagaacagggggccagaactctcggcagtaagaataaaaggccagacagagaggc
agcagcacatatctgcttcgcacacagctgcaatcactagcaagctctcaggcctggcatcatggtgcattttact
gctgaggagaaggctgcgctcactagcctgtggagcaagatgaatgtggaagaggctggagggtgaagccttgggca
55 ggtgaagcattgggtctcaatgcattgggaatgaagggtgaatatattaccctagcaagttgattgggaaagtccccaag
attttttgcactctcaattttgtatctgatatgggtgcattttcctagactcctcgttgtttaccctctggaccaga
gaatttttgacagcttggaaaacctgtcgtctcctctgcactcctgggcaaccctcaaggccaagggccatggcaca
gaaggtgctgactctcctttgagatgctattaaaaacatggacaaacctcagcccgctttgtctaaqctgagtga

ctgcactgtgacaagctgcatgtggatcctgagaacttcaaggtgagttcaggtgctggtgatgtgatttttggc
 tttatatatttgacattaattgaagctcataatcttattggaaagaccaacaagatctcagaaatcatgggtcgag
 cttgatgttagaacagcagacttctagtgcagcataacccaaacttacatgattcagaactagtgcagtaaaaggac
 tactaacagcctgaattggcttaacttttcaggaaactttgccagaacttgatgtgtttatcccagagaattgtat
 5 tatagaattttagactttgtgaaagaagaatgaaatttgcttttggtagatgaaagtcattttcaaggaaatagaa
 atgccttattttatgtgggtcatgataattgaggtttagaaagagatttttgcaaaaaaataaaagatttgctca
 aagaaaaataagacacattttctaaaatatgttaattttcccatcagtatgtgaccaagtgaaggcttggttccg
 aatttgttggggatttttaactcccgtgagaactcttgacgactcacattctacattttacaaaaattagacaat
 tgcttaagaaaaacaggagagagggaacccaataatactggtlaaaatggggaaggggtgagggtgtaggtagg
 10 tagaatgttgatgtagggtcatagaataaaattgaacctaaagctcatctgaatttttgggtgggcacaaacct
 tggacagtttgaggtcaggttgcttaggaatgtaggatataaagccgtttttgtttgtttgttttttcatc
 aagttgttttcggaaacttctactcaacatgcctgtgtgttattttgtcttttgccaaacagctcctgggtaacgt
 gatggtgattattctggctactcactttggcaaggagttcaccocctgaagtgcaggctgctggcagaagctgtg
 tctgtgtgcacattgcctggcccataagtaaccactgagttctctccagtttgagggtgttctgtgacctga
 caccctccttctgcacatggggactgggttggtcttgagagaaagccttctgtttaataaagtacattttcttca
 15 gtaatacaaaattgcaattttatcttctccatcttttactcttggtttaaaggaaaaagtggtcatgggtgagg
 gatggagagaaacataggaagaaccaagagcttccctaaagaaatgtatgggggttgtaaaattaatgtggatgtt
 atggggagaattccaggattccaaggaggatgatgtatggagaaaaatctttatcggggtgggaaaaatggttaatt
 aagtgacagagactcctaggcagtttttactgcaccggggaagaaggagctgttagtggtacctgagaaagcag
 atttgggtacatgtcacttttcaataaaaaacaaaaacaaaaacaaacttcatagatatccaagatatagg
 ctagaattactattttaatttactcttatttacttttggaagttagctagcttgccatgtttttagaaattgatt
 20 tggagataagatgagtggtgtatcaacaatagcctgctcttccatgaaggattccattatttcatgggttagctga
 agctaagacacatgatatactgtgtcattatcttctgatagaatgtaacatgcactaaaaataaagtttagagttagg
 acctgagtgaggaaagtttttgagagtgatgaagactttccgtgggagatagaataactaataaaggcttaaat
 ctaaaaccagcaagctagggtctcgtagcttgcatgaaactggctctctggaagtagaaggagagtaagacatac
 gtagaggactagaaagaccagatagtagcagggtcgtgtacaaaaatacaagcttttactatgctatttgcaatac
 25 taaacgataagcattaggtatgttaagtgcactcaggaaataagattttgggaaaaagtaactcgtctatgtgcacaa
 aatggattcaagtttgagataaaaaataaataatggatgatgattcaaggggacagatacaatgggtcaaacccaag
 aggagcagtgagctgtggaatttgaaagatggacaaaggtgggtgagaaagacatagatttcgactgactgtg
 gagatgagaagggaaggaaggaggtgataaatgactgaaagctcccagactggtgaagataacaggaggaaacctgc
 actgacctgggtgactctcatgtgtgaagggttagagggatattaacagatttacttttttaggaagtgtgattgggt
 caggaggttttgaccttcaggtcttgtgtctttcatatcaaggaaacctttgcattttccaagtttagagtgccatat
 30 tttggcaaatataacttttattagtaattttatagtgctctcacattgatcagacttttctgtgtaattacttttg
 aatttggctgtatataccagaatatgggagagagacaaataattattgtagttgcaggctatcaacaatactggt
 ctctctgagccttataacctttcaatatgcccataaacagagtaaacagggtatttcatggcactaaatattttc
 acctagtcagtcacaaatgggagcaatgtgcatttttgatacatatttttatatatttttaggggtacatgtgat
 actttacatgctagaacatgtgatgattaagtctagatattttaagatatccattgctttgagcatttatctttct
 35 atgtattgagaaaaattcaaatcctcatttctagccattttgaaatatataataaataagtaatttaactatagtcac
 cctactcaaatatcaaacattatggcttaactcttctatccaaactgtgtttgtacctattaaaccaacatctcttaa
 atccctcccatcacactcacactttttccagcctctgataactatcattctactctctaccaccatgagaccca
 ctttttttagctcccacagatgaataaaaaacatgtgataatttgactttctgtatctggctatttttatctatct
 ctttggcataccaagagtttgtttttgttctgcttcagggtcttcaattaacataaatgacctctggttccatccat
 40 gttgtacaaatgacaagatttcattcttttcatggcaaaatagtagctgtgcaaaaatacaatttttaattccgt
 tcatctgttgatagacacttaggttgatcccaaaccttaacatttgtaaatagtgcttcaataaacatgagtgtaa
 tgtgtccattggaataactgatttcccttcttttggtataaataaccactagtgagattgctggattgtatgtagt
 tctgttttttagtttactgagaaatcttcatactgttttccataatgggtgtactattttacattccaccaacagt
 gtgtaagaaagagttcccttttctccatctcctcacaggatctgttattttttgtctttttgttaatagccgtt
 ttaactagagtaagtagatatctcattgtagttttgatttgcatttccctgatcattagtgatgttgagaattttt
 45 tcatatgtttgttgggtcatttgtatatcttttctgagaattgtctgttcatgtccttagccctactttttattggg
 attgtttgttattttcttgataatctatttgtgttcattttagagcctggatatttcttttgtcagatgtatag
 attgtgaagattttctccactctgtgggtgtctgtttattctgcagactcttcccttttgccatgcaaaagctct
 tttagtttaatttagtcccagatatattttctgttttttagtatttgcatttgtgttcttgggtcatgaaatccttct
 ctaagccaatgtgtagaagggtttttccgatgttattttctagaattgttacagtttcagggtttagatttaagtc
 cttgatccatcttgagttgattttttgtataaggtgagagatgaagatccagtttcattctcctacatgtagcttgc
 50 cagctatccccgcaccatttgttgaatagggtgccctttcccaactttatgtttttgtttgttcaagatca
 gttggatgtaagtatttgagtttatttctgggtctctattctgttccattgggtcgatgtgcctatttgtacacca
 gcatcatgctgttttgggtgactatggcctatttgtatagtttgaaatgaggtaaatgtaatgccttcagatttgttc
 ttttttttagacttgctgttttatttgggtctctttttgggtccataagaatttttaggattgtttttctagttctg
 tgaagactaatgggtgatttttgatgggaattgcaatgaattttagggtgtcttctggcattatggccattttcac
 aatattgattctacccatctatgagaatggcatgttttccatttgtttgtgtcttattatgattactttcagccgt
 55 gttttgtagttttccttgtagatgtcttccactccttgggttaggtatataatccaaagtttttgttttgtttgt
 tttgtttttgcagctattgtaaaaggggttaggttcttgtattttattctcagcttgggtcattgtcgtgtatgtaag
 aaagcaactcattgggtgtacgttaattttgtatccagaaactttgctgaattattttatcagttctagggggtttt

ggaggaggtcttttagagttttctacatacacaaatcatatcatcagcaaacagtgacagtttgactttctctttaaca
 atttggatgtgctttacttgtttctctgtctgattgctcttgctaggacttccagtaatatgttaaagagaagtg
 gtgagagtggggtatccttgtctcattccagttttcagacagaatgcttttaacttttccattccaatataatgtt
 ggcgtgtgttttaccatagctggccttttattacattgaggtatgtcctttgtaaaccgatttttgctgagtttttagt
 cataaagtgtatgtgaattttgttgaaatgcagtttctgtggctatttgagataaatcacatgatttttgtttccaatt
 5 ctcctttatgttgtgtatcacacttattgacttgcgtatgttaaaccatccgtgcatccctcgcatgaaacccactt
 gatcatgggttttgatatgctgtcggatgctattagctagtattttgtcaaggatgttggcatctatgttcatcag
 ggatattgatctgtagtgttttttttttttgggtatgttctttccagttttgggtattaaagtgatactggcttca
 tagaatgatttagggaggattctctcttctctatctttagaataactgtcaataggattgggtatcaattcttctt
 10 tgaatgtctgtgagaattcagctgtgaatctatctggctcctggacttttttgttgttggtaaatttttattatcat
 ttcagttctgtgcttattactggctgttccagggatctaatcttctcctgacttaagctagagccctgtatcttt
 ccaggaattcgaacgtctccttttaggttttctagtattatgcatgtaaagggtttcatagtagccttgaataatctt
 ttgtatttctgtgtgtatcagtaatagtatctcctgttttggtttctaattgagtttatttgcacttctcctctt
 tcttgggttaattctgtcaatgggtctatcagttttatttattcttttcaaagaaccagctttttatttcatttagctt
 15 ttgtatttttttgcagttgttttaatttcatttagttctcctcttatcttagttattccctttcttttgtctgggtt
 ttgggtctgtttgtttttgtttctctagttttcttgggtgtgaccttatattgtctgtctgtcctctttcagactc
 tttgacatcgacatttagggctgtgaactttccttttagcaccatcttctgtgtatcctagagggttttgatagggt
 gtgtcactattgtcggtcagttcaagtaattttgttgtcttattataactttaagttctgggatacatgtgcagaa
 tgtgcagggtttgtacataggtatagatgtgccatgggtgggttgcctgcacccatcaacctgtcatctacattaggt
 atttcttttaattgttatccctctcctaaccctccaccccccagcagggcctgggtgtgtatgttccctccctgt
 20 gtccatgtgttctcattgttcaactccacttatgagtgagaaacgtgtgggtgtttgggttctctgttccctgtgtta
 gtttgtctcagaatgatgtgttccaccttcatccatgtccctgcgaagacatgaaactcatcttttatggctgcat
 agtattccatgggtgtatattgtgcacattttctttatccatttatctcgtgatggccatttgggttgggttccaagt
 ctttgtctattgtgaatagtgtccacaataaacatacgtgtgcacgtgtctttatagtagaatgatttctaattcttt
 ggggtatatacccagtaatgggattgtctgggtcaaacagttattctgggtctatagatccttgagggaatcgccacactg
 tcttccacaatgggtgaactaatttacacacccatcaacagtgtaaaatttttctatcttccacatcctctcca
 25 gcaccttttgtttcctgactttttaataattggcatttctaactggcatgagatgggtatctcatttgggttttgatt
 tgcatttctctaataagcagtgatgatgagcttcttttcatgtgtttcttggccacataaatgacttcttttagaga
 agcatctgttccatctcctttgtccactttttgatggggtcggttaggtttttcttgaatttttggtaagttcttt
 gtagattttggatgttagcccttgtcagatggatagattgcaaaaattttctccattctgttaggttgcctgttc
 actctgatgatagcttttctgtgtgcagaagctcttttagtttaattagatcccataatgtcaattttggcctttgtt
 gtcattgtcttttgatgttttagtctgtgaatttttggccatgcctatgtcctgaatgggtattgcctaggttatcttc
 30 taggatttttatgggttttaggttgcacatttaagtccttaattccaccttgagtttaattttgtataaggtgttaagg
 aaggggtacagttttaggttttatgcataattgctagccagtttttccagcaccatttattaaatagggaaattcttct
 tccattgtcttttgtgatgtttgtcaaatgcagatgggtcgtatagtggtggcatttatttctgaggcttctgttctg
 ttccactggcttatatctgttttggtaaccagtaaccatgctgttttggtaactgtagcctttagatagtttttga
 agtcaggtagcatcatgcctccagctttgttcttttggtttaggttgccttggctatatgggctcttttttgatt
 ccataatgacatttaagtagttttttctaattctttgaaaaaagtcagtggttagcttgatggggatagcattgaat
 35 ctataaattactttgggcagtatggccattttaagatatgtattctttctatctatgagcatggaaatgtttttcc
 atttgtttgtgtcctctcttatttcccttgagcagtgagtggtttgttagctctccttgaagaggttcttcacatcc
 ttagaagttgtatttctaggtattttattttattctcttgcagcaattgtgaatgggagttccaccatgatttgg
 ctctctgcttgtctatttattgggtgtataggaacgcttgtgatttctgcacactgattttgtatcttgagactttgc
 tgaagctgtttatcagcttaagatttttgggtgagatgacagggcttcttaaatatacaatcatgtcatctgcaaa
 40 cagagacaatttgacttccctctcttcttatttgaattatgtcttatttcttctcttgcctgatttgcctggcgaga
 acttccaataactatgttgagtaagagtggtgagagggcactccttgtcttgtgcccgttttcaaagcaaatgatttt
 taaatttccatcttgatttcatgtttagcccaatgatcattcaggagcaggttattttaaatttccctgtatttgc
 ggttttgaaggttcccttttaggttgaatttccaattttattctactgtggtctgagagagtgcttgatataatttc
 aatttttaaaaaatttattgaggttgttttgtggcatatcatatggcctatcttggagaaagttccatgtgtgat
 gaatagaatgtgtattctgcagttgttgggtagaatgtcctgtaaatatctgttaagttccatttgttctttaaatc
 45 cattgtttctttagtagactgtcttgatgacctgectagtgagtcagtgaggtattgaagtccccactattatta
 tgttgcctgtctagcttagtagtaattgttttataaatttgggtatctccagttattagatgcataatattaagaatt
 gtaattattctccatttgacaagggttttattcattatatagtgctcctcttgtcttttttaactgctgtttctt
 taaagtttgttttgtctgacataagaatagctgttggctcgttttgggtgtccatttgtgtggaaatgtcatttt
 ccacccctttaccttaagtttatgtgagtccttatgtgttaggtgagtcctgaaggcggcagataactgggttgg
 tgaattcttattcattctgcaattctgtatcttttaagtggagcatttagtccatttacattcaacatcagtttg
 50 aggtgtgaggtactattccattcttctgtgttatttggctgtgtatctttttatctgtatttttgttgtatag
 tccatagggatttatgctttaaagaggttctgttttgatgtgtctccaggggtttatttcaagatttagagctcctt
 ttatcagttctttagtagtttggcttggtagtgccgaattctctcagcatttgtttttctgaaaaaactgtgtatt
 ttcttctatttgaagcttagtttactggatataaaattcttggctgataattgttttgttttaagaaggctgaag
 ataggccatattcacttctagcttttaccggttttctgtgagaaatctgctgtttaaactgtataggttttcttcaat
 55 aggttacctggtagtttccctcacagctcttaagattctcttgtcttttagataaacttggatactctgatgaca
 atgtacctaggcaatgatatttttgcattgaatttccaggtgtttatttagacttcttgtatttggatattctaggt
 ctctagcaagggtgggggaagttttccttgattatttccctggataagttttccaaacttttagatttctcttctt

ctcaggaatgctgattattcttaggtttgattgttttaacataatcccagattttcttgaggctttgttcatat
 cttattcttttttctttgtctttgttggattgggttaattcaaaaactttgtcttcaagctctgaatttcttctgc
 ttggattctattgctgagactttctagagcattttgcatttctataagtgcattccattcatccattgtttctctgaa
 gttttgaaatgtttttatttatgctatctctttaactgaagatttctccctcatttcttgtatcatatttttgggt
 5 ttttttaaaattggacttccacttccctcggaatctatttgattagcttaataactgaacttctgaatttttctc
 aggtaaatcagggatttcttcttgggtttggatgcattgctggtgagctagtagattttttgggggtgttaaaga
 acctgtttttcatattaccagagtttagtttctgggttcttctcacttgggtaggctctgtcagagggaaagtct
 aggcctcaaggctgagacttttgtcccatgaggtgttcccttgatgtagcacagtcccccttttctaggcgtggg
 gcttctgagagccgaactgtagtattgttatctctcttctggatctagccaccatcagggtctaccagactcca
 10 ggctggtactgggtttgtctgcacagagtcttctgacgtgaaccatctgtgggtctctcagccatagatacaacc
 acctgctccaatggaggtggcagaggatgaaatggactctgtgagggctccttacttttgggtgttcaatgcactat
 ttttgtgctggttggcctcctgcccagaggtggcactttctagaagcatcagcagaggcagtcaggtgggtgggtgg
 ctgggggggctggggcaccctagaactcccaagaatatattgccccttgtctcagctaccagggtagtaaggaa
 gaccatcaggtgggggcaggactagtcgtgtctgagctcagagtctccttgggcaggtcttctgtggctactgtg
 15 ggaggatgggggtgttagtttccagggtcaatggatttatgttcttaggacaattatggctgectctgctgtgtcatg
 caggtcatcaggaaagtgggggaaagcaagcagtcacgtgacttggccagctcccatgcaactcaaaaggttggct
 tcacttccagcgtgcaccctccccgcacagcaccgaatctgtttccatgcagtcagtgagcaaggctgagaact
 tgccccaggtaccagctgcgaaaccaagtagggctgtcctacttccctgcccagtgaggctgtgcacaccaaattca
 tgtccccccaccaacccccccactgcccagcccctagatctggccaggtggagattttcttttctgtcatcttt
 tccaggttctctggcagccctcccaatgacccctgtgaggcagggcagaatggcttcttaggggaaccagaga
 gccacagggcttttcccgctgcttccctctacccctgtattttgttggccctctaaaattgactcagctccagta
 20 aggtcagaatcttctcctgtggtctagatcttcaggttccccagtgaggatgtgtgtttgggggtagacgggtcccc
 cttttccacttccacagtttgggcactcacaatatttgggggtgtttcccggtcctgcaggagcaatctgcttctt
 tcagaggggtgtgtgcgttctctcagcttcttctgatttatttctgcaggtggttctgcacaaaaaattctgatggg
 agacttcacatgctgctctgtgcacccagtgaggctgcaatgtacttctgctgectcccatctgcatcaccct
 ctaatttgtcggtaatatgcatttttaataatcttttttctctctctctcttttcttctccccaaaaactata
 25 ctgccccttgatatcaaggaaatcaaggacgtgatgttgaggggtgggcagtgagatacactctttacccttaggga
 gctatatctagatttagatatgtccaattcaagataacttaattgaaagcaaatccataatgaatacacacacaca
 cacacacatctgcatgacaagatttttaatagttgaaagaataactaataattgtccacaggcaataagggtttt
 taagcaaaacagttgtgataaacaggtcattcttagaatagtataatccagccaatagtacaggttgccttagagatta
 tgtcattaccagagttaaaaattctataatggcttctcactccctaccactgaggacaagtttatgtccttaggttt
 atgcttccctgaaacaataaccactgctatttctccactttacatatcaacggcactggttcttcttaactctct
 30 ggcacagcaggagtttgttttcttctgcttcagagcttgaatttactatttcagcttctaaactttatttggcaa
 tgccctcccatggcagattccttctgtcattttgcccctgttgcgaatactttctccttaatttcattcttagttaa
 taatatctgaaattatttttgttgtttaacttaattatttaatttatgtatgttctaccttagattataatcttcaga
 ggaagtttttatttctgacttatttaacttaaatgcccactctttaaataattatgacatttttaacacagatat
 ttgttgaacaaatgtttgaaaatacatgggaagaatgtcctgaaaacacttgaaattgcttgtgtaaagaacagt
 35 tttatcagtttaggatttaataatgtcagaagcaatgatataaggaaaaatcgagggaataagacagttatggataag
 gagaaatcaacaaactctttaaagataattgcccataaaagcataagaggaaataagggtttatacatgacttttaga
 aactgccccttgggttttggataaatggggaagttgtttgaaaacaggagggatcctagatattccttagtctgagg
 aggagcaatgaagattcacttgttttagaggtgggaagtggtggtcagcctgtaatccagaaatttgggagggcc
 aaggcaggcagatcacctgaggtcaagagttcaagaccaactggccaacatggtgaaatcccatctctacaaaaa
 taaaaaatttagacaggcatgatggcaagtgcctgtaatccagctacttgggaggtgaggaaggagaattgtctt
 40 gaacctggaaggcaggagttgcagtgagccgagatcataccactgcactccagcctgggtgacagaacaagactct
 gtctcaaaaaaaagagagattcaaaagattcacttgttttaggccttagcgggttagacaccagctctctgac
 acattcttaaaggtcaggctctacaaatggaaccaaccagactctcagatatggccaagatctatacacacca
 tctcacagatccccatctttaaagagaccctaatttgggttccctcagctctctataatctgtaccagcataccaa
 taaaaatctttctacccatccttagattgagagaagtcattattattatgtgagtaactggaagatactgataa
 45 gttgcaaaatcttttcttcttcttattcaacttttattttaacttccaaagaacaagtgcattatgtgcagc
 tttgttggcaggtcaacatgtatcttctgtgtcttttagccgctaactttgagcagatataagccttacaca
 ggattatgaagtctgaaaggattccaccaatattattataattcctatcaacctgataggttaggggaaggtagag
 ctctcctccaataagccagatttccagagtttctgacgtcataatctaccaaggtcatggatcgagttcagagaaa
 aaacaaaagcaaaaccaaactcaccaaaaataaaaaatcccaagaaaaataaagaaaaaacagcatgaatact
 tctgcccattgttaagtggccaatatgtcagaacagcactgagttacagataaagatgtctaaactacagtgacat
 50 cccagctgtcacagtggtggaactatttagtcaataaaacagtcctgccccttaagagttgttttccatgcaata
 catgtcttatgtcttagaataagattccctaagaagtgaacctagcattttatacaagataatttaattctaataccat
 agtatctggttaaagagcattctaccatcatctttaccgagcatagaagagctacacaaaaacccctgggtcatcagc
 cagcacatacacttatccagtgataaatacacatcatcggtgctacatacatacctgaatataaaaaaataact
 tttgtgagatgaaacaggcgtgatttttcaaataggtagcgataagtagatatgaaagtaaggattcagtcctt
 atattatattacataacatttaattctctgacgtgaaactgtgtctttataggatttttctactacactacagag
 55 aacttaagagataatggcctaaaaccacagagagtagtattcaagataagtagacttcttatttggaaaccaa
 tgcttactaaatgagactaagcgtgtcccatcaaaaatcctggacctatgcccataaacacatttcacaatccctg
 aacttttcaaaaattggtacatgctttaactttaactacaggcctcactggagctacagacaagaaggtagaaaaa

5

10

15

20

25

30

35

40

45

50

55

gagactaagattttgtccgctcaaaaatcctggacctatgctcaaaacacatttccacaatccctgaacttttcaaaa
atttggtacatgcttttagcttttaactacaggcctcactggagctagagacaagaaggtaaaaaacggctgacaaaa
gaagtcctggtatcctctatgatgggagaaggaaactagctaaagggaagaataaatttagagaaaaactggaatga
ctgaatcggaaacaggcaaggctataaaaaaaattagcagtatcctcttgggggccccttcccacactatctca
5 atgcaaatatctgtctgaaacggtccttggtctaaactccacccatgggttggccagccttgcccttgaccaatagcc
ttgacaaggcaaaacttgaccaatagctcttagagtatccagtgaggccaggggccggcggtggttagggatgaaga
ataaaagggaagcacccttcagcaggttccacacactcgcttctggaacgtctgaggttatcaataagctcctagctcc
agacgccattgggtcatttcacagaggagacaaggctactatcaacagcctgtggggcaaggtggaattgtggaagat
gctggaggagaaaacccctgggaaggtaggctctgggtgaccaggacaagggaggggaaggagaccctgtgcttggca
aaagtccagggtcgcttctcaggattttgtggcacccttctgactgtcaaacgttcttctgcaatctcacaggctcctg
10 gttgtctacccatggaccagaggttctttgacagctttggcaacctgtcctctgacctgacctcatgggcaacc
ccaaagtcaaggcacatggcaagaagggtgctgacttcttgggagatgccacaaagcactggatgatctcaaggg
cacccttggccagctgagtgaactgcactgtgacaagctgcatgtggatcctgagaacttcaaggtagtccagga
gatgtttcagccctgtgctttagctctcgaggcaacttagacaacggagatttgatctgagcacagcagggtgtg
agctgtttgaagatactgggttgggggtgaagaaactgcagaggcataactgggtgagaccagtggttaagtgtt
ttagggtctaaggagtgccttcaaaaactagatggacaattttagctttgagaaaagagaggttggaattgagaa
15 aatgactttttctttattagattccagtagaagaactttcatctttccctcatttttgttgttttaaacatctat
ctggaggcaggacaagtatggctggttaaaaagatgcaggcagaaggcataatattggctcagtc aaagtggggaaact
ttggtggccaaacatacattgctaaggctattcctatatcagctggacacatataaaaatgctgctaagtcttcatt
acaaacttatatcctttaattccagatgggggcaagtatgtccaggggtgaggaacaattgaaaacatttgggctg
gagtagattttgaaagtccagctctgtgtgtgtgtgtgtgtgtgcgcgcgcgcgtgtgtgtgtgtgtgtgtgtgtgt
20 gtttcttttaacgtcttcagcctacaacatacagggttcatggtggcaagaagatagcaagatttaaatatggcc
agtgcattgtgcttgaaggggaaacactacctgcatttaattggaaggcaaaatctcaggtctttgagggaaagttaa
cctaggcttgattctgggtgggaagccttggtgtgtagtattctggaggccaggtggagctctcagctcactatggg
ttcatctttattgtctcctttcatctcaacagctcctgggaaatgtgctggtgaccgttttggcaatccatttccg
caaagaattcacccctgaggtgcaggcttctctggcagaagatggtgactgcagtggccagtgccctgtcctccaga
25 taccactgagctcactgcccattgattcagagctttcaaggataggtttattctgcaagcaatacaaaataataaat
ctattctgtctgagagatcacacatgattttcttcagctcttttttttacatcttttttaaatatagagccacaaag
ggtttatatgtgagggaaagtgtgtatgtgtattttctgcatgctctgtttgtgtttgtgtgtgtgtgtgtgtgtgtgt
tatttttatatgagatgtgtcatttttagtgagcaaaataaaagcagtaaaagacacttgatcacggaggttctgcaagt
gggagtaaatgggtgtaggagaabaatccgggtgggaagaaagacctataggacaggacttctcagaaacagatgtttt
30 tgggaagagatggggaagggttcagtgaagacctggggctggttattgattgcagctgagtagcaaggatggttctta
aggaagggaaagtgttccaagcttttaggaattcaagggttttagtcagggtgtagcaattctattttattaggaggaat
actattttctaattggcactttagcttttcacagcccttgtggatgcctaagaaagtgaattaatcccatgccctcaa
gtgtgcagattgggtcacagcatttcaaggagagagacctcattgtgaagactctgggggaggtggggacttaggtgta
agaaatgaatcagcagaggctcacaagtcagcatgagcatgttatgtctgagaaacagaccagcactgtgagatca
aatgtagtgggaagaatttgtacaacattaatgggaaggcttacttaattggaatttttgtatagttggatgttag
35 tgcactctataagtaagagtttaatatgatgggtgtacggacctaatgtttgtgtcctcctcaaaattcacatgct
gaatccccactccccactgaccttatctgtggggaggtctttgaaaagtaattaggttttagatgagctcaatgaag
agcagatccccatcataaaattattttccttatcagaagcagagagacaagccatttctcttctcctccgggtgagg
acacagtgagaagtccgccatctgcaatccaggaagagaacctgaccacagagtcagccttcagaaaatgtgagaaa
aaactctgttgttgaagccaccagtccttttgtattttgttatagcacttgactgagtaaggcagatgaagaag
40 gagaaaaaaataagcttgggttttgagtggactacagaccatgtttatctcaggtttgcaaagctcccctcgctcc
ctatgtttcagttataaaaatccctactctactactctcatctataagacccaataataagcctgcgccttctctc
taactttgattttctctatttttacttcaacatgctttactctagccttgtaattgtctttacatacagtgaaatgt
aaagttctttattctttctttctttctttctttctcctcagcctcagaatttggccacatgccttctctctct
caggaaacttctccaacatctctgcttggctccatcatatacaaaaggtccacttcaaatgcagtactactcgtttt
cagaatatgcactttctttcttttttgtttttttttaaagtc aaagcaaatcttcttgagagagtaaaagaaa
45 taaacgaatgactactgcataggcagagcagccccgaggggccgtggttgttcccttttatggttatttcttgatga
tatgttaaacaaagttttggattattttatgccttctcttttttaggccatatagggttaactttctgacattgccatgg
catttttcttttaatttaatttactgtttaccttaattcagggtgtacacgtacaggatgtgcaggtttgttttata
ggtaaaagtgtgccatggttttaattgggtttttttttcttgttaaggtgttttaagtttcttgtttactctggata
ttaggccctttgtcagaagaatagattggaaaatcttttcccatctctgtagattgtctttcgctctgatggttagtt
tcttttctgagcagagctcttttagtttaatttagattccatttggtcaatttttctgtcttttctgtcaattgtctttc
50 acgtttctcatcatgaaatctgtgcctgttttatatcatgaatagattatgcttggatttttcttaggctttttat
agtttgggttttttcttttaagtctctaattccatctggagtttaatttgggataagggtataagggaaggagtccagtt
tcatttttcagcatatggctagccagttcttccccatcatatttaaaattgaaaatcccttccccattgtctgtct
ttgtcagggtttctaaaagaccagatgggtgtagggtacaatatgcagtttcttcaagtcataataaccatctgaaa
tctcttatttaattcatttctttttagtatgtatgctggtctcctctgctcactatagtgagggcaccattagccaga
gaatctgtctgtctagtctcatgtaagattctcagaattaaagaaaaatggatggcatatgaatgaaactctcatggat
55 gacatatggaatctaataatgtattgttgaattaatgcataagatgcaacagagagaaggtgacaactgcaatgat
aacctggattgatgatataagagtctatagatcacagtagaagcaataatcatggaacaaatgtgaaattgggga
acagccacaaacaagaagaatacaatacttccaggaagaagtgcagtcaggtcacttttctggagcgggtgagagaa

aagtggaagttagcagtaactgctgaattcctgggttgctgatggaaagatggggcagctgttactggtacgcag
 ggttttagatgtatgtacctaaggatagaggtatggcaatgaacagaaattcttttgggaatgagtttagggcc
 attaaaggacatgacctgaagtttctctgaggccagtcacacaaactcaatataaatgtgttctctgcatatagt
 caaagttgcccacttcttttcttcatatcatcgatctctgctcttaaagataatcttgggttttgcctcaaactggt
 5 tgtcactacaaaactttcccatggttcttaagtaaaacaggtaactgcctctcaactatacaagtagactaaaata
 ttgtgtctctaatatcagaaattcagctttaatatattgggtttaactctttgaaatttagagctctccttgaaata
 cacatgggggtgatttctctaaactttatttcttgtaaggatttatctcaggggtaacacacaaaccagcatcctga
 acctctaagtatgaggacagtaagccttaagaatataaaataaactggttctctctctgcccgtggaagtgtgccc
 tgtctatttctgaaattgcttgtttgagacgcatgagacgtgcagcacatgagacacgtgcagcagcctgtggaat
 10 attgtcagtgagaatgtcttctgctgattagatataaagacaagttaaaccacagcatagactatagatcaagcc
 tgtgccagacacaaatgacctaatgcccagcagggccacgggaatctctctctctgcttgaacagagcagcac
 acttctcccccaacactattagatggttctggcataattttgtagatatgtaggatttgacatggactattgttcaa
 tgattcagaggaatctccttggttcagataagtaactgactactaaatggattaaaaaacacagtaataaaaacc
 cagttttcccttacttccctagtttgttcttattctgcttcttccaagttgatgctggatagaggtgtttatt
 15 tctattctaaaaagtgtgaaattggccggggcgcggtggctcacacctgtaatcccagcactttgggaggctgagg
 tggcgagatcacgaggtcaggagatcaagaccatcctggctaacatgggtgaacccccatctctactaaaaatacaa
 aaaattagccagagacagtgccgggtgctgtagtcccagctactcgggaggctgaggcaggagaatggcggtgaac
 ctgggaggcagagcttgccgtgagcagagatcgcgccactgcacactccagcctgggtgacaaagcgagactccat
 ctcaaaaaaaaaaaaaaaaaaagaaaaagaaagaaagaaaaaactgatgaaattgtgtattcaatgta
 gctcacaagaaattgaaaaccaagaaagctgtggcttctccacataaagcctggatgaataacaggataaacag
 20 ttgttacattgtcacaaactcctgctcaggaattgagtggtcaagatattcgttaattcttactcctttcagttgtaa
 cttattcctatttgtcagcatcagggttattagcggtgctggcgaagtccttgagaataaaactgcacactggat
 ggtgggggtagtgtaggaaaaatggaggggaaggaagtaaaagtttcaaattaaagcctgaacagcaaagttccctga
 gaaggccacctggattctatcagaaactcgaatgtccatcttgcaaaacttcttgcccaacccccacccctggag
 tcacaacccacccttgaccaatagattcattttactgagggaggcaagggtggtcaatagattcatttctactgg
 gagaggcaaaagggtgggggcccagagaggagaagtaaaaagccacacatgaagcagcaatgcaggcatgcttctgg
 25 ctcatctgtgatcaccaggaaactcccagatctgacactgtagtgcatttctactgctgacaagaaggctgctgcca
 ccagcctgtgaagcaaggttaaggtgagaaggctggaggtgagattctgggcaggtaggtactggaagccgggaca
 aggtgcagaaaggcagaaagtgttctgaaagagggttagcccggtgtcttacatagctgacttgcacctgct
 ctgtgatattgactatcccacagctctcctgggtgtctaccttaggacctagaggtaactttgaaagtttggatct
 tgggctctgactgtgcaataatgggcaacccccaaagtaaggccacatggcaagaaggtgctgatctccttcggaaa
 agctgttatgtctcaggtgacctcaaaggcacttctgtacactgagtgacctgcactgtaacaagctgcacgtg
 30 gacctgagaactcctggtgagtagtaagtaactcagccttcttctttacccttagatatattgcaactatgggt
 acttttgaaagcagaggtggttctcttctgtgttatgagtcagctatgggatatgatatttcagcagtggtgattt
 gagagttatgttctgttaataacataactaaaaatttggtagagcaaggactatgaataatggaaggccacttacc
 attgatagctctgaaaaacacatcttataaaaaattctggccaaaatcaaaactgagtgtttttgatgagggaaac
 agaagttgagatagagaaaaataacatcttcttgggtcagcgaaatttctataaaaaatgaatgactcttct
 35 gcatagtcctggaggttagaaaaagatcaactgaacaaagtagtggaagctgttaaaaagaggattgtttccctc
 cgaatgatgatggtatacttttgtacgcatggtacaggattcttgttatgagtggttgggaaaattgtatgtatg
 tatgtatgtatgtatgtatgactggggacttatcctatccattactgttcttgaagtaactattatcctactttt
 taaaaggacgaagtctctaaaaaaaatgaacaaatcacaatatgttggggtagtgagttggcatagcaagtaag
 agaaggataggacacaatgggaggtgcagggtgcagctcatattgaagctgatattctagcccataatgggtgagag
 40 ttgctcaaactctggtgaaaaaggtatgaaggttatatactatttactgcaagtcagcttgaggccttctattca
 ctatgaccatttcttttttacttctcactccctcccgagctcttaggcaacgtgatattgattgttggcaaac
 cacttcagcgaggtattttaccctacagatacaggtcttcttggcagtaactaacaaatgctgtggttaatgctgtag
 cccacaagaccactgagttccctgtccactatgtttgtacctatggtccactatgtttgtacctatgtcccaaat
 ctcatctccttttagatgggggaggttggggagaagagcagtatcctgctgctgattcagttcctgcatgataaaa
 atagaataaagaatatgtctcttaagaaatatcatgtactcttttctgtcttttatattttaccctgattcagc
 45 caaaaggacgcactatttctgatggaaatgagaatgttggagaatgggagtttaaggacagagaagatacttctt
 gcaatcctgcaagaaaagagagaactcgtgggtggtatttagtggggtagttactcctaggaaggggaaaatcgtctc
 tagaataagacaatgtttttacagaaaggaggtcaatggaggtactcttggagggtgaagaggattgttggtag
 tgtgtagaggtatgttaggactcaaattagaagttctgtataggtattattttgtatgaaactcaggatatagctc
 atttgggtgactgcagttcacttctacttatttttaaacacatatatttttattatttataatgaagtggggatgggg
 50 cttcttagagaccaatcaagggccaaacctgaaactttctttaaactcctcaatgggtattaatagagaattatct
 ctaaggcatgtgaaactggtgtcttgggttttcatctgtacttctcatctgctacctctgtgacctgaaacatat
 aattccattaagctgtgcatatgatagatttatcatatgtattttccttaaaggatttttgaagaactaattgaa
 ttgatacctgtaaaagtctttatcacactacccaataaataaataatctcttctgtcagctctctgtttctataaat
 atgtacaagttttattgttttttagtggttagtgattttattctcttctatataatatacacacacatgtgtgcat
 55 ataaatatatacaattttttatgaataaaaaattatttagcaatcaatattgaaaaccactgatttttggttatgtga
 gcaaacagcagatataaaagctgagatttaggaacagcagcttaagtcaagttgtatagaggagaatggacatt
 taaaagaggcaggtgatataaaatttagggaaactggatgcagagaccagatgaagtaagaaaaatagctatcgtt
 ttgagcaaaaactcactgaagtttcttgcatatgagagtgacataataaataagggaacgtagaaaattgattcaca
 tgtatatatatatatagaactgattagacaaagtctaacttgggtatagtcagaggagctgtctgtaatttatattg

aggtgatggataaagaactgaagttgatggaaacaatgaagttaagaaaaaaaaatcgagtaagagaccattgtggc
 agtgattgcacagaactggaaaacatttgtgaaacagagagtcagagatgcagagctaaaatccctgtctgtgaatga
 aaagaaggaaatttattgacagaacagcaaatgctacaagcccccgtttggatctggcaatgaacgtagccatt
 ctgtggcaatcacttcaaactcctgtacccaagacccttaggaagtatgtagcaccctcaaaccctaaaacctcaaa
 gaaagagggttttagaagataataaccccttcttccagtttcattaatcccaaacctcttctcaaagtattt
 cctctatgtgtccaccccaagagctcacctcaccatactcttctgagtgaggagcacatagataggcggtgtacca
 tctaacagcttctgaaattcctttgtcatatttttgagtccccactaataaccacaaagcagaataaataccagt
 tgctcatgtacaataatcactcaactgctgtctttagtcatacattaattaagcacattctttgaataattactgt
 gtccaaacaatcacacttttaaaatctcacactttagtcatcccttgcccttctgaatgtcactctgtattttaaat
 gaagagatgagggttgaatttctgtgttacttattgtttcatttctcgatgaggagttttcacattcacctttagt
 ggaaaacacataagtagacacatcttacaggaaaaatataccaactgacatgtagcatgaatgcttgtgcatgtagt
 catataaaattctgtagcaatgtaaacattctctgatatacacatacagatgtgtctatatgtctacacaatttct
 tatgtctccatgaacaaacattccatgcacacataagaacacacactgtttacagatgcataacttgagtgcaattgaca
 aaattaccccagtcaatctagagaatttggatttctgcatttgcactctgttagctttagcatgtgttcatttac
 tctgggtgatgtcttccctcattttgccttgtctatcttgtactcatactttaagtccctaaacttatatgtttatct
 caactaagaagctattttttttaatttttaactgggcttaagccctgtctataaaactctgctacaattatgggct
 ctttcttataatatttagtgttttccctactaatgtacttaattctgtctattgttatctcctaccactaaatttta
 acctcttttatggtagagacattgtcttgtaaactcttatttccctagtatttggagatgaaaaaaagattaaat
 tatccaaaattagatctctcttttctacattatgagtattacactatccatagagaagtttggtttgagacctaacc
 tgaggaccttttggttctaaaatgactatgtgatctcttagtatttataggtcatgaggttccctcctctgtcctct
 gctatagtttgattagtcacaagcatgtgtcatgcatttattcacatcagaatttcatacactaaatagacatag
 tatcagaagtcagtttattagttatatcagttagggtccatcaaggaaaggacaaaccattatcagttactcaacc
 tagaattaaatacagctcttaatagttaattatccttgtatttgaagagctaaaatatcaaataaaggacagtgca
 gaaatctagatgttagtaacatcagaaaacctcttccgccatttaggcctagaagggcagagaggagaaaatgtttat
 accaccagagtcagagaaccagagcccataaccagaggtccactggattcagtgagctagtgggtgtccttggaga
 gagccagaactgtctaattgggggcatcaaggtatcagccataaaaaaccataaaaaagactgtctgtctgtaggaga
 tccgttcagagagagagagagaccagaaataactcttgccttatgctttccctcagccagtggtttaccattgcagaat
 gtacatgcgactgaaagggtagaggaaacctgggaaatgtcagttcctcaaatacagagaaactgaggggaaggatg
 agaaataaatgtgaaagcagacatgaatggtaattgacagaaggaaactaggatgtgtccagtaaatgaataatta
 cagtggtgcagtgattattgcaatgattaatgtattgataagataatatgaaaacacagaattcaaacagcagtgaa
 ctgagattagaattgtggagagcactggcatttaagaatgtcacacttagaatgtgtctcttaggcattgttctgtg
 catatatcatctcaatattcattatctgaaaattatgaattaggtacaaagctcaaataaatttatttttcaggtt
 agcaagaactttttttttttttctgagatagagcattgtctatgggtgccaggctggagtgcaatggcatgatc
 caggctcactgcaacatctgctcccaggttcaagcgattctcctgctcagccctcccaagtagctggcactacag
 gcatgtgccaccaccatgctggcctaattttctatttttagtagataggggggtttccaccatgttgggtcaggctgat
 ctgcaactcctaatactcaggtgatccaccctcctggcctctgaaagtgtctgggatcacaggcgtaggccaccaca
 cccagccaagaattgtgaattttagtagaaggatataaacocatatttctctgacctagagtocttagtataacctccc
 ataccatgtggctcatcctccttacatacatattcccatctttcacccctaccttttcttttggtttcagcttttca
 ctgtgtcaaaatctagaaccttatctcctacctgtctgaaaccaacagcaagttgacttccattctaaaccacat
 tggcattacactaattaaaatcgatactgagttctaaaatcatcggggattttggggactatgtcttacttcatac
 ttcccttgagatttcacattaaatgttgggtgttcattaaaggtccttcatttaactttgtattcatcacactcttgg
 attcagagttatatctaaactcttaatacagcctgtataatcccaattcccaactctgattttctaaaccttgacc
 tccaacctcagtgccaaaccatataatcaacaatgtactgggcttattttatatagatgtcctataggaacctcag
 actcagcatgggtatttccactgtttataactaaaactgtttctctccagtggttttccatttttagtcatagatagc
 tacttgccatttcaccaaggtcacagattaaaatcatttccctacctctaataacagttcgattctgtcttcaatt
 tgtccctatctattaatcaccaactcttactgccagtcaggtcctcattgttttctgaacaagagtagatgtctatt
 ctttccacttttagaccttatcctggctggatgcggtggctcaggttgttaaaccagcactttggggaggccaagg
 caggcagatcacttgagggtcaggagttcaagaccagcctgaccaacatggtgaaaccccatctctactaaaaatcac
 aaaatcagccggcggtgtgtgtgcatgctgcagtcacagctattcaggtggctgaggcaggagaattgtctgaacc
 caggaggcagaggttgggtgagcctagattgcaccattgcactctagcttgggcaatagggatgaaactccatct
 cagaagagaaaaaagaaacacttattctgttatcaaaactcctctcaatgcaatccatatagaataaacatgta
 accagatctcccaatgtgtaaaatcatttcaggtagacagaaataaagtgaaggaagcagctcttgggaattaaaca
 gacaaagatcaaaataacagtcctcatggccttaagaatttacctaaccatttttttagaatacaatttcttatata
 tgaattggaaacataattcctccctcacaaacacattctaaagattttaaggagatattgatgaagtacatcatctg
 tcatttttaacaggttagtggttagtgattcacacagcacattatgatctgttcttgtatgttctgttccattctgta
 ttcttgacctggttgtattcttctgtagctccagatccacatataagtagacatctttttgcattttacaagagtg
 catacaatacaatgtatccaagactgtatttctgattttatctgtaccactaaaactcacaaatgtggccctattctt
 gtgttcaagactgacatcacctcatggtccaaagtctgataatagaaatggcattgtcactttcttccctactgca
 acagaagcccagctatttgtctccattttctctacttctaaaaatacatttcttccactaagtgagaataatctttt
 aaagacacaaatcaaacatgccaccacttcttggtaatttcaatatcttcttgggtgcttccaggttacagaaaa
 aataacttgtaacaaagtttaaaaggtcattcatggctcctctctaccctattttataaacatttcccttgtgatca
 gaatctcaggcacatcatccatttctatatatacaataaagtcataatagtttgaactcacctctgggttactttta
 atcaaccaaatgctgtaaaatgcatttgtatcgctacgtgttaagcagtagttgattcttttcttctgtgtaat

attctattctttgactataaccgtaatttatcaattctactgttggttaagcatttaagtggctaccggtttgaggtt
 tttatgattattgtctgtcataagcattttctatacatgtctttggatacacacatgcattgtttctgaatatctaa
 aaatgtaattgtctaggttaataagacttatcaagcatccagcattttgtggatactattaaaggttttccaaaggggtt
 atactattgtacagtgtcaccacacagagtttgagtttctattgatccatataccaccacaaaatttgaactgtcagt
 5 cttatctcttctctgtctcttttttctctcttttttctcttcccttccctctctctctctctctctctctctct
 tctctctcttct
 ct
 ttct
 cattgtggcattatttgaattcagaaaagagtgaaaactactgggatcttcatctctgggtctaatccacatttt
 10 ttttttaagaacacatctgtaaaaatgttctgtactagcatattcccagggaacttcgttaaaatcttggtga
 atatggtaaatctacttttcaactttgcattctttcttttagtcataaccataattttaaacattccaaaatattgtat
 ataataatttgattttatctgtcattaaaaatgttaaccttaaaatctcatgtttccagaacctatttcaataactgg
 aaataaacactattcatttttttaaatattcttttaattggatatttatttcaatataataaaaaaattagagtttat
 tataggaagaatttacaaaaagaaggaggaagcaagcaagtttaaaactgcagcaatagatttgtccattccaacct
 ctcaaaattcccttgagacaaaaatctctagaggcaagaagaactttatattgagtcaacttggttaaacatct
 15 gcttttagataagttttcttagtataaaagtgcagaaacaaataagttaaactctaagatacattccactatatta
 gcctaaaacacttctgcataaatgaaactaggaggtatatttttagaaacaaactgctgaaagagatgcggtggggag
 atatgttagaggagaacagggtttctgagtcagacacacatgcagaaacagccaatctcagggcaagtttaagggaa
 tagtggaatgaaggttcatttttctctctcacaactaatgaaacctgcttatcttaaaccaacctgctcactgg
 agcagggaggacaggaccagcataaaaaggcagggcagagtcgactgttgcttacactttctctgacataacagtgt
 20 ttcaactagcaacctcaaacagacacccatgggtgcatctgactctgaggagaagactgctgtcaatgccctgtgggg
 caaagtgaacgtggatgcagttgggtgggtgaggccctgggcaggttggtatcaagggtataagagaggctcaaggag
 gcaaatggaaactgggcatgtgtagacagagaagactcttggtttctgataggcactgactctctgtcccttggtg
 ctgttttctctacctcagattactgggtgtctaccttggtgaggggtcttttgagtccttggtggatctgtcc
 tctctgatgctgttatgggcaacctaaaggtgaaggtctatggcaagaaggtgctaggtgcctttagtgatggcc
 25 tgggtcacctggacaacctcaagggcactttttctcagctgagtgagctgactgtgacaagctgcacgtggatcc
 tgagaacttcaggtgagtcaggagatgcttcaacttttctcttttacttttcaatcttacatttttggttctttt
 acctacctgctctctctccacatttttgctattttactatattttatcatttaagtctctcaaaattttgttaatt
 ttttatttaaatattctgcatttttctctctcacaactctgctatttttaaatatttaatatctctgtctttctc
 tcccaacccccctcccttcatttttctctctctcaacaacaaactcaaatatgcataccagctctcacctgctaattc
 30 tgcacttagaataatccttttctctctccacatgggtatgggagagggtccaaactcaaatgagagggcatagaat
 actgttttagaggctataaatcattttacaataaggaataatgggaattttataaattctgtagtaaatggaatgg
 aaaggaaagtgaatatttgatattgaaagactaggcagttacactggaggtggggcagaagtcgttgctaggagac
 agcccatcatcactgattaatcaattaatattgtatctattaatctgtttatagtaattaatattgtatagctat
 atacacatacaaaaataaaaactaatttggaattaatattgtatagttatatacagcatatatagcataatagtac
 35 atatatagactacatgctagttaagtacatagaggatgtgtgtgtatagatatagtttatagtagcattcatat
 atgtacttatttatgctgagtggaataaacctggggaatcagttttgtctaaagatttgggcagaaaaaatgggtgt
 ggctcagtttctcagaagccagtcctttatttctctgttaaccatagcatgtatctgcctacctctctctccgcagc
 tcttgggcaatgtgctggtgtgtgtgctggcccgcaactttggcaagggaattcacccacaaatgcagggtgccta
 tcagaaggtggtggtggtggtgctaatgcccgtgctcacaagtaaccattgagatcctggactgtttctgataac
 40 cataagaagacctatttccctagattctattttctgaacttggaacacaaatgcctacttcaagggtatggcttc
 tgcttaataaagaatgttcagctcaacttctgattaatttcaacttatttctattttttgtccaggtgtgtaagaa
 ggttctgaggtctacagataggagcacttggtttattttacaagagtagcatgggaaaagagaaaagcaaggga
 accgtataaggcattaatgggtgacacttctacctcaaaagagcagaaattatcaagaactcttgatacaaaagata
 atactggcactgcagaggttctaggggaagacctcaacctaaagacatagcctcaagggttaagtctacagattaaact
 ccaacaattactgagaaaataatgtgtcaattaaaggcataatgattactcaagacaatgttatgttgtctttct
 45 tctctctcttcttgctgcacattgtagcccataatactataccccatcaagtgcttctgctccaagaaatagctt
 cctctcttacttgccccagacatctctgtaaagaatttctcttctatcttcccatatttcaagtaagattcattg
 ctacgtattacttgtgacctctcttgacccagccacaataaacttctctatactacccaaaaaatctttccaaa
 cctccccccacaccattttttatatttttatatttttctattttatctcatgcacacacacactccgtgcttta
 taagcaattctgctattctctaccttcttacatgctactgtgctcatattaaattcatcaatgggcagaaaga
 50 aaatattttattcaagaaaacagtgaatgaatgaacgaatgagtaaatgagtaaatgaaggaatgattatctctgc
 tttagaacttctggaattagaggacaatattaataataccatcgacagtggtttcttctgttgtaagtctacaaca
 tacaagagggaagcatgcagtaaacacccgaacagttatttctctctgatcataggagtaatattttttctcttg
 agcaccatttttgccataggtaaaattagaaggatttttagaactttctcagttgtatacatttttaaaatctgt
 atttatgcatgttgattaattttaacttacttgaatacctaaacagaatctgttggttctctgtgttgaaagt
 55 gctttcacagtaactctgtctgtactgccagaatatactgacaatgtgttatagtttaactgttttgatcacacat
 tttgaattgactggcagcagaagctcttttatatccatgtgttttcttaagtcattatacatagtaggcactgag
 aactctttatctgaataagatatttaggaaccactggtttacatatcagaagcagagctactcagggcattttg
 gggagagcaccttccattctctgagcatagggaggttctcataagagtaagatattaaaaggagatacttgtgtg
 gtattcgaaagacagtaagagagattgtagacctatgatcttgatagggaacaaactacattctcttctccaa
 aagtcaaaaaaaagagcaaatatagcttactataccttctattctctacaccattagaagtagtcagtgagctag
 gcaagatgttgccctaaaaatccaaataaccagagaattcatgagaacatcacctggatgggacatgtgccgagca

5

10

15

20

25

30

35

40

45

50

55

aattgttttcttttgtttaattotttgcctttctttttttttcttctcgcgaatttttactattatacttaatgcctt
 aacattgtgtataacaaaaggaaatatctctgagatacattaagtaacttaaaaaaaaaactttacacagctctgcct
 agtacattactattttggaatataatgtgtgcttattttgcatattcataatctccctactttattttcttttattttt
 aattgatacataatcattatacatatttatgggttaaagtgtaatgttttaatatgtgtacacataattgaccaa
 cagggttaattttgcattttgtaatttttaaaaaatgctttcttcttttaatatactttttgtttatcttattttctaa
 tactttccctaactctctttctttcagggcaataatgatatacatgtatcatgcctctttgcaccattctaaagaata
 acagtgataattttctgggttaaggcaatagcaatatctctgcatataaaatattttctgcatataaattgtaactgat
 gtaagagggtttcatattgctaataagcagctacaatccagctaccattctgcttttattttatgggtgggataaggc
 tggattattctgagtcgaagctaggcccttttgcataatcatgttccatacctcttatcttctccacagctcctgg
 gcaacgtgctggtctgtgtgctggcccatcacttttggcaagaattcaccaccagctgcaggtgcctatcagaa
 agtgggtggctggtgtggctaatgccttggccacaaagtatcactaagctcgctttcttgcctgctcaattttctatta
 aaggttccctttgttccctaagtcactactaaactgggggatattatgaagggccttgagcattcggattctgcgc
 taataaaaaacattttattttcattgcaatgatgtattttaaattatttctgaaatattttactaaaaaggggaatgtg
 gaggtcagtgcattttaaacataaagaatgaagagctagttcaaaccttgggaaaatacactatatcttaaactc
 catgaaagaagggtgaggtgcaaacagctaatgcacattggcaacagccctgatgcataatgccttattcatccct
 cagaaaaggattcaagtagaggttgatttggagggttaaagttttgctatgctgtattttacattacttattgttt
 tagctgtcctcatgaatgtcttttactaccatttgccttatcctgcatctctcagccttgactccactcagttct
 cttgcttagagataccacctttccctgaagtggttctctccatgttttacggcgagatgggtttctcctcgccctggc
 cactcagccttagttgtctctgttgtctttagaggtctacttgaagaaggaaaaacaggggtcatgggttgactg
 tctgtgagcccttcttccctgctcccccactcagctgacccggaatctgcagtgctagtctcccggaactatc
 actctttcacagctctgctttggaaggactgggcttagtatgaaaagttaggactgagaagaatttgaaaggcgct
 tttttagacttgatattcactactgtcttattacccgtgcataggcccccaccccaaatggaagtccattcttctctc
 aggatgtttaagatttagcattcaggaagagatcagagggtctgctggctcccttatcatgtcccttatgggtgcttct
 ggctctgcagttatttagcatagtgttaccatcaaccaccttaacttctattttcttattcaatacctaggtaggta
 gatgctagattctggaaataaaaatgatgtctcaagtggtccttgcctctctccagtcacaaattctgaatctagt
 tggcaagattctgaaatcaaggcatataatcagtaataagtgatgatagaagggtatataagaagaattttattata
 tgagagggtgaaacccctcaaaatgaaatgaaatcagacccttgccttacaccataaaacaaaaataaatttgaatgg
 gttaaagaatttaactaagacctaaaaccataaaaatttttaagaagaatcaaaagaagaaaatttctaataattcacg
 ttgcagccggtttttgaatttgatgtagaagcaaaaggcaacaaaaggaaaaataaagaagtggaggtacatcaaa
 ctaaaaaatttccacacaaaaacaaaaaatgaacaaatgaaagggtgaaccatgaaatggcatatttgcacaacca
 aatattttcttaaatatttttggttaatatccaaaatatataagaacacagatgattcaataacaaacaaaaatta
 aaaataggaaaaataaaaaatataaaagaagaaaatcctgccatttatggcagaattgatgaacctggaggatgta
 aaactaagaaaaataagcctgacacaaaaagacaaaatactacacaaccttgcctcatatgtgaaacataaaaaagtc
 actctcatggaaacagacagtagaggtatgggtttccagggggttgggggtgggagaatcaggaaactattactcaa
 gggataaaaatttcagttatgtgggatgaataaattctagatatctaattgtacagcatcgtgactgtagttaattg
 tactgtagatatatttaaaatttgcaaaagagtagatttttttttttttttagatggagttttgctcttgttgtc
 cagggtggagtgcaatggcgaagatcttggctcacttgcaacctccgctcctgggttcaagcaaatctcctgctca
 gctcccgagtagctgggattacaggcatgagacaccatgcccagctaattttgtatttttagtagagacgggggtt
 tctccatgttggctcaggctgatccgctgcctcgccaccccaagggtcgggattacaggcgtgagccaccgggccc
 tggccgagagtagatctttaaagcatttaccacaagaaaaaggttaactatgtgagataatgggtatgttaattagc
 ttgattgtggtaatcatttcacaaggtatacatataataaaacatcatgttgtaacacctaaatatatacaatttt
 tattttgtgaatgatacctcaataaagttgaagaataataaaaaagaatagacatcacatgaattaaaaaactaaaa
 aataaaaaaatgcacttgatgattagaattgcattcttgatttttcagatacaaatatccatttgactgtttact
 cttttccaaaacaatacaataaaatttttagcactttatcttcttcttcccttcccaatctataattatataatat
 atatttttagatattttgtatagtttttactccctagatttttctagtgttattattaaatagtgaagaatgtttaca
 cttatgtacaaaatgttttgcagctttttcttcttcttaacattctctcttaagttttattctattttttctgatt
 atccttaatatattctcttctgctggaatacattgttacttttgggtttatctaaaaatggcttcattttcttca
 ttctaaaatcatgttaaatataaccactcatgtgtaagtaagtagtggaataaatagaatccaaaaactaaat
 ctactaaaatataataatgtgatataaaaaatataagcttttaaaatttagcttggaaataaaaaacaaacagtaa
 ttgaacaactatactttttgaaaagagtaaaagtgaatgcttaactgcataataccacaatcgattacacaattagg
 tgtgaaggtaaaattcagtcacgaaaaaactagaataaaaaataggggaagacatgtatataatcttagagataaca
 ctgttattttaattatcaaccacaaagtagaaactatcaaggagaaataaattcagtcacaataaaaagcatttaag
 aagttattctagctgggagcggtggctcacacctgcaattgcagcactttgggaggcctagacagcggtacag
 acgtcaggagttcaagatcagcctggccaacatagtgaacctcatcgctactaaaaataaaaaacttagcctgg
 cgtggtggcaggcatgtgtaateccagcaatttgggaggctgaggcaggagaatcgcttgatcctgggaggcagag
 gttgcagtgagccaagatttgccactgcattccagcccaggtgacagcatgagactccgtcacaaaaaaaaga
 aaaaaaaagggggggggggagcggtggagccaagatgaccgaataggaacagctccagctctatagctcccatcgtg
 agtgacgcagaagacgggtgatttctgcatttccaactgaggtaccaggttcatctcagagggaagtgcaggcag
 tgggtgcaggacagtaggtgcagtgcaatgtgcatgagccaaagcagggcgaggcatcacctcaccgggaagcac
 aagggtcaggggaattcccttcttagtcaaaagaaaggtgacagatggcactggaaactcgggtcactcccgcc
 cctaatactgcgctcttccacaagcttaacaaatggcacaccaggagattatatcccatgcctggctcagagggt
 cctacgcccatggagcctcgctcattgtctagcacagcagctctgaggtcaaactgcaaggtggcagtgaggctgggg
 gagggtgcccaccattgtccaggcttgagcaggtaaacaaagccgctggaagctcgaaactgggtggagcccacc

acagctcaaggaggcctgctgctctgttaggctccacctctaggggcagggcacagacaaacaaagacaacaag
 aacctctgcagacttaaatgtccctgtctgacagctttgaagagagtagtggttctcccgacacatagcttcagat
 ctgagaacaggcagactgctcctcaagtgggtccctgacccccgagtagcctaactgggaggcatccccagtag
 gggcagactgacacctcacatggctggtactcctctaagacaaaacttccagaggaatgatcaggcagcagcattt
 5 gcggttcaccaatataccactgttctgacgccaccgctgttgatacccgagaaaacagcttctggatggacctcca
 gtaaaactccaacagacctgacgtgagggtcctgactgttgaaggaaaactaacaacagaaaggacatccacac
 caaaaacccatctgtacatcgccatcatcaaagaccaaaggtagataaaaccataaagatggggaaaaagcagagc
 agaaaaactggacactctaaaaatgagagtgcctctcctcctccaaagtaacgcagctcctcaccagcaatggac
 aaagctgggcagagaatgactttgacgagttgagagaggaaggcttcagaagatcaaactactccaagctaaagga
 10 ggaagttcgaacaaacggcaagaagtaaaaaactttgaaaaaaattagatgaatggataactagaataaccaat
 gcacagaagtcttaaaaggacctgatggagctgaaaaccaaggcaggagaactacgtgacaaatcacaaagcctca
 gtaaccgatgagatcaactggaagaaagggtatcaatgacgaagatgaaatgaatgaaatgaagcatgaagagaa
 gtttagagaaaaaagaataaaaaagaaacgaacaaagcctccaagaaatattgggactatgtgaaaagaccaatcta
 catctaattgggtgtagctgaaagtgatggggagaatggaaaccaagtggaaaaactctgcaggatattatccagg
 agaacttcccaatctagcaaggcaagcccaaatctacattcaggaaatacagagaacgccacaaagatactccta
 15 gagaaaagcaactccaagacacataactgtcagattcaccaaagttgaaatgaaggaaaaaatgttaagggcagcc
 agagagaaaggtcggttaccacaaagggaagcccatcagactaacagctgatctatcggcagaaactctacaag
 ccagaagaaagtgggggccaataattcaacattgtttaagaaaaagaaatttcaaccagaaattctatatccagccaa
 actaagcttcataagtgaaggagaaataaaaatcctttacagacaaagcaaatgctgagagattttgtcaccaccagg
 cctgcccataagagctcctgaaggagcactaaacatggaaaggaacaactagtatcagccactgcaaaaacatg
 20 ccaaatgttaagaccatcaaggctaggaagaaactgcataacagagcaaaaataaccagctaacatcataatgaca
 ggatcaaattcatacataacaataactcaccttaaatgtaaataggctaaatgctccaattaaaagacacagactgg
 caaattggataaggagtcagaccccatctgtgttctgtatttcaggaaacccatctcacgtgcagagacacacatag
 gctcgaaataaaaaggatggaggaatatctaccaagcaaatggaaaacaaaaaaaggcaggggttgcaatcctagt
 tctgataaaacagatttttaaccaacaaagatcaaaagagacaaagaaggccattacataatggcaagggtatcta
 ttcaagaagaagaactaactataactaatatataatgcacccaatacaggagcaccagattcataaaacaagtcct
 25 gagtgcctacaagagacttagatgccacacaataataatgggagactttaacaccccactgtcaacattagac
 agatcaacgagacagaaagttaacaaggatatccaggaattggactcagctctgcaccaagcagacctaatagaca
 tctacagaactctccaccccaaatcaacagaaatatatacttctttcagcaccacaccacacctattccaaaactga
 ccacatagtttgaagttaaagctctcctcagcaaatgtaaaagaacagaaactataacaaactgtctctcagaccac
 agtgcaatcaaactagaactcaggattaaagaaactcactcaaaacccactcagctacatggaaactgaacagcctgc
 30 tccatgaatgactactgggtacataacaaaatgaaggcagaaataaagatgttctttgaaaccaacgagaacaaaga
 cacaacacaccagaatctctgagacacattcaaaagcagtggtgtagagggaaatttatagcactaaatgccacaaag
 ggaaagcaggaaagatctaaaattgacacccataacatcacatttaaaaaactagagaagcaggagcaaacacattc
 aaaagctaacagaagacagaataaactaagatcagagcagaagtgaaggacatagagacacaaaaaaccccttca
 aaaaaatcaatgaatccagaagctgttttttgaaaagatcaacaaaattgatagactgtagcaagactaataaa
 35 gaagaaaagagagaagaatcaaatagacgcaataaaaaaatgacacgggtatcacccactgtatccacagaaataca
 aactaccgtcagagaataactataaacacctctacgcaataaactagaaaaatctagaagaatggataaattctct
 gacacatacactctgccaagactaaaccaggaagaagttgtatctctgaaatagaccaataacaggctctgaaattg
 aggcaataattaatagcttatcaacaaaaaaagtcgggaccagtaggattcatagccgaattctaccagaggta
 caaggaggagctggtaccattctctgaaactattccaatcaatagaaaaagagggaatctccttaactcattt
 40 tatgaggccagcatcatctgataccaaagcctgacagagacacacaaaaaaagagaatgttacaccaatatcct
 tgcataacattgatgcaaaaaactcctaataaataactggcaactgatccaccatgatcaagtggtggttcatccct
 gcatgcaaggctggttcaacatacgaataatcaataataatccagcatataaacagaaaccaaagacacaaacc
 atatgattatctcaatagatgcagaaaaggcctttgacaaaattcaacaacgcttcatgtctaaaaactctcaataa
 attaggtattgatgggacatatctcaaaataataagagctatctatgacaaacccacagccaatatcatactgagt
 45 ggacaaaaactggaagcattccctttgaaaactggcacaaggcagggtatgcctctctcaccactcctattcaaca
 tagtgttgtaagttctggccagggaatcaggcaggagaaggaataaagggcattcaattaggaaaagagggaagt
 gaaattgtccctgtttgcagatgacatgattgtatatctagaaaacccattgtctcagcccaaaatctccttaag
 ctgataagcaacttcagcaaaagtctcaggatataaaaatcagtggtgcaaaaatcacaagtatcctatgcaccaata
 acagacaaacagagagccaaatcatgagtgaactccattcacaaattgcttcaagagaataaaaatcaccttaggaat
 50 ccaacttacaaggatgtgaaggacctctcaaggagaactacaaacccactgctcaatgaaataaaaagaggatataca
 acaaatggaagaacattccatgctcatgggtaggagaatcaatatcgtgaaaatggctcatactgcccagggttaa
 tttatagattcaatgccatccccatcaagctaccaatgactttcttcacagaactggaaaaaactactttaaagtt
 catatggaacaaaaaagagcccacataccaaggcaatcctaagccaaaagaacaaagctggaggcatcacgcta
 cctgacttcaactatactacaatgctacggtaaccaaacagcatggtactggtaccaaaacagagatctagacc
 aatggaacagaaacagagccctcagaaataatgcgcataatctcaactatctgatctttgacaaacctgagagaaa
 caagcaatggggaaaggattccctatttaataaaatgggtgctgggaaaactggctagccatatgtagaagctgaaa
 55 ctggatcccttcccttacaccttatacaaaaatttaattcaagatggattaaagacttacatgttagacctaaaacca
 taaaaacccctagaaaaaaaacctaggaataccattcaggacatagggcatgggcaaggacttcatgtctaaaacacc
 aaaagcaatggcaacaaaagacaaaatggacaaacgggactcaattaaactaaagagcttctgcacagctaaagaa
 actaccatcagagtgaacaggcaacctacaaaatgggagaaaatttttgcaatctactcatctgacaaagggttaa
 tatccagaatctacaatgaactcaaacaaatttacaagaaaaaaacaaacaccccatcaaaaagtgggcaaggat

atgaacagacacttcgcaaaagaagacatttatgtaatcaaaaaacacatgaaaaaatgctcatcatcactagcca
 tcagagaaatgcaaatcaaaaccacaatgagataccatctcacaccagttagaatggcgatcattaaaaagtcagg
 aaacaacaggtgctggagaggatgtgggagaacagggaacaacttttacactgttgggtgggactgtaaactagttca
 accattgcggaagtcagtggtggaattcctcagggaatctagaactagaaataccatttgaccagccatcccatta
 ctgggtacatacccaaaggattataaatcatgctgctataaagacacatgcacacgtatgtttattgcagcactat
 tcacaatagcaaaagacttggaaaccaacccaaatgtccaacaacgatagactggattaagaaaatgtggcacatata
 cccatggaatactatgcagccataaaaaatgatgagttcatgtcctttgtagggacatggatgaagctggaaact
 atcattctcagcaaaactatcacaaggagaataaaccaaacacccgcatgttctcactcataggtgggaattgaacaa
 tgagaacacatggacacatgaagaggaacatcacactctggggactgttatgggggtggggggcaggggcagggata
 gcactaggagatatacctaatagtctaaatgacgagttaatgggtgcagcacaccaacatggcacatgtatacatata
 taacaaacctgcatgttgtgcatgttaccctaaaacttgaagtataataataaaaaaagttatcctattaaaac
 tgatctcacacatccgtagagccattatcaagtcttctctttgaaatagacagaaatttagtgttttctcagtca
 gttaac [SEQ ID NO:19]

[0081] Five 5' hypersensitive site (HS) sites (HS1-HS5) and one 3' HS site have been identified in the human β -globin LCR (Stamatoyannopoulos *et al.*, (2001)). The 5' HSs 1-4 are Dnase I hypersensitive sites. The HS2 and HS3 elements are the most powerful single elements within the LCR (Ellis *et al.*, EMBO J. (1996), 15:562-568; Collis *et al.*, EMBO J. (1990) 9:233-240), as corroborated by many groups. Deleting HS2 in the context of β YAC in transgenic mice severely affects HS site formation as well as expression of all of the human β -globin genes at every developmental stage (Bungert *et al.*, Mol. Cell Biol. (1999);19:3062-3072). It was reported that deletion of HS2 minimally reduced the expression of the embryonic ϵ y and β hi globin genes in yolk sac-derived erythrocytes (Ley *et al.*, Ann. N.Y. Acad. Sci. (1998);850:45-53; Hug *et al.*, Mol. Cell Biol. (1996);26:2906-2912). HS2 functions primarily as an enhancer.

[0082] In certain embodiments, the β -globin LCR region comprises a HS2 region. In non-limiting example, the β -globin LCR region comprises a HS2 region, a HS3 region, and a HS4 region. In certain embodiments, the HS2 region, HS3 region and HS4 region within the β -globin LCR region are contiguous. In one non-limiting embodiment, the β -globin LCR region consisting essentially of a HS2 region, a HS3 region and a HS4 region. In another embodiment, the β -globin LCR region comprises two introduced GATA-1 binding sites at the junction between the HS3 region and the HS4 region. The HS3 region can lie between the HS2 region and the HS4 region. The length and the sequence of the HS2 region can vary. The HS2 region can have a length of from about 400 bp to about 1000 bp, *e.g.*, from about 400 bp to about 500 bp, from about 500 bp to about 600 bp, from about 600 bp to about 700 bp, from about 700 bp to about 800 bp, from about 800 bp to about 900 bp, or from about 900 bp to about 1000 bp. In one non-limiting embodiment, the HS2 region has a length of 860 bp. In one non-limiting example, the HS2 region has the nucleotide sequence set forth in SEQ ID NO:9, which is provided below:

GTATATGTGTATATATATATATATATATATTCAGGAAATAATATATTCTAGAAATATGTCA
 CATTCTGTCTCAGGCATCCATTTTCTTTATGATGCCGTTTGAGGTGGAGTTTGTAGTCAGG
 TGGTCAGCTTCTCCTTTTTTTTGGCCATCTGCCCTGTAAGCATCCTGCTGGGGACCCAGAT
 AGGAGTCATCACTCTAGGCTGAGAACATCTGGGCACACACCCTAAGCCTCAGCATGACTC
 ATCATGACTCAGCATTGCTGTGCTTGAGCCAGAAGGTTTGCTTAGAAGGTTACACAGAAC
 CAGAAGGCGGGGTGGGGCACTGACCCGACAGGGGCTGGCCAGAACTGCTCATGCTTG
 GACTATGGGAGGTCACTAATGGAGACACACAGAAATGTAACAGGAACTAAGGAAAACTG
 AAGCTTATTTAATCAGAGATGAGATGCTGGAAGGGATAGAGGGAGCTGAGCTTGTAATA
 AGTATAGTAATCATTAGCAAAATGGTTTTGAAGCACCTGCTGGATGCTAAACACTATTTT
 CAGTGCTTGAATCATAAATAAGAATAAAACATGTATCTTATCCCCACAAGAGTCCAAGT
 AAAAAATAACAGTTAATTATAATGTGCTCTGTCCCCAGGCTGGAGTGCAGTGGCAGCAT
 CTCAGCTCACTGCAACCTCCGCCTCCCGGGTTCAAGCAATTCTCCTGCCTCAGCCACCCT
 AATAGCTGGGATTACAGGTGCACACCACCATGCCAGGCTAATTTTTGTACTTTTTGTAGA
 GGCAGGGTATCACCATGTTGTCCAAGATGGTCTTGAAGTCTGAGCTCCAAGCAGTCCAC
 CCACCTCAGCTCCCAAAGTGCT [SEQ ID NO:9]

[0083] In certain embodiments, the HS2 region has a length of about 840 bp. In certain embodiments, the HS2 region has a length of about 650bp (*e.g.*, 646 bp). In certain embodiments, the HS2 region has a length of about 420bp (*e.g.*, 423 bp).

[0084] The length and the sequence of the HS3 region can vary. The HS3 region can have a length of from about 200 bp to about 1400 bp, *e.g.*, from about 200 bp to about 300 bp, from about 300 bp to about 400 bp, from about 400 bp to about 500 bp, from about 500 bp to about 600 bp, from about 600 bp to about 700 bp, from about 700 bp to about 800 bp, from about 800 bp to about 900 bp, from about 900 bp to about 1000 bp, from about 1000 bp to about 1100 bp, from about 1100 bp to about 1200 bp, from about 1200 bp to about 1300 bp, or from about 1300 bp to about 1400 bp.

EP 3 620 521 A1

In certain embodiments, the HS3 region has a length of about 1300 bp. In one non-limiting embodiment, the HS3 region has a length of 1308 bp. In one non-limiting embodiment, the HS3 region has a length of 1301 bp. In one non-limiting example, the HS3 region has the nucleotide sequence set forth in SEQ ID NO:5, which is provided below:

```
5      AAGCTTTCATTAAAAAAGTCTAACCAGCTGCATTTCGACTTTGACTGCAGCAGCTGGTTA
      GAAGGTTCTACTGGAGGAGGGTCCCAGCCATTGCTAAATTAACATCAGGCTCTGAGACT
      GGCAGTATATCTCTAACAGTGGTTGATGCTATCTTCTGGAACCTGCCTGCTACATTGAGA
      CCACTGACCCATACATAGGAAGCCCATAGCTCTGTCTGAACTGTTAGGCCACTGGTCCA
10     GAGAGTGTGCATCTCCTTTGATCCTCATAATAACCCCTATGAGATAGACACAATTATTACT
      CTTACTTTTATAGATGATGATCCTGAAAACATAGGAGTCAAGGCACTTGCCCCCTAGCTGGG
      GGTATAGGGGAGCAGTCCCATGTAGTAGTAGAATGAAAAATGCTGCTATGCTGTGCCTCC
      CCCACCTTTCCCATGTCTGCCCTCTACTCATGGTCTATCTCTCCTGGCTCCTGGGAGTCA
      TGGACTCCACCCAGCACCACCAACCTGACCTAACCACTATCTGAGCCTGCCAGCCTATA
15     ACCCATCTGGGCCCTGATAGCTGGTGGCCAGCCCTGACCCCACCCCACCCCTCCCTGGAAC
      CTCTGATAGACACATCTGGCACACCAGCTCGCAAAGTCACCGTGAGGGTCTTGTTGTTGC
      TGAGTCAAAATTCCTTGAAATCCAAGTCCTTAGAGACTCCTGCTCCCAAATTTACAGTCA
      TAGACTTCTTCATGGCTGTCTCCTTTATCCACAGAATGATTCTTTGCTTCATTGCCCCA
      TCCATCTGATCCTCCTCATCAGTGCAGCACAGGGCCCATGAGCAGTAGCTGCAGAGTCTC
20     ACATAGGTCTGGCACTGCCTCTGACATGTCCGACCTTAGGCAAATGCTTGACTCTTCTGA

      GCTCAGTCTTGTCATGGCAAAATAAAGATAATAATAGTGTTTTTTTATGGAGTTAGCGTG
      AGGATGGAAAACAATAGCAAAATTGATTAGACTATAAAAGGTCTCAACAAATAGTAGTAG
      ATTTTATCATCCATTAATCCTTCCCTCTCCTCTCTTACTCATCCCATCACGTATGCCTCT
25     TAATTTTCCCTTACCTATAATAAGAGTTATTCTCTTATTATATTCTTCTTATAGTGATT
      CTGGATATTAAAGTGGGAATGAGGGGCAGGCCACTAACGAAGAAGATGTTTCTCAAAGAA
      GCCATTCTCCCCACATAGATCATCTCAGCAGGGTTCAGGAAGATAAAGGAGGATCAAGGT
      CGAAGGTAGGAAC TAAGGAAGAACA CTGGGCAAGTGGATCC [SEQ ID NO:5]
```

30 **[0085]** In certain embodiments, the HS3 region has a length of about 850 bp (*e.g.*, 845bp). In certain embodiments, the HS3 region has a length of from about 280 bp to about 290 bp (*e.g.*, 280bp and 287bp).

[0086] Similarly, the length and the sequence of the HS4 region can vary. The HS4 region can have a length of from about 200 bp to about 1200 bp, *e.g.*, from about 200 bp to about 300 bp, from about 300 bp to about 400 bp, from about 400 bp to about 500 bp, from about 500 bp to about 600 bp, from about 600 bp to about 700 bp, from about 700 bp to about 800 bp, from about 800 bp to about 900 bp, from about 900 bp to about 1000 bp, from about 1000 bp to about 1100 bp, or from about 1100 bp to about 1200 bp.

35 **[0087]** In certain embodiments, the HS4 region has a length of about 1.0 kb or more. In certain embodiments, the HS4 region has a length of about 1.1 kb. In certain embodiments, the HS4 region has a length of about 1150 bp (*e.g.*, 1153 bp). In one non-limiting embodiment, the HS4 region has a length of 1065 bp. In one non-limiting example, the HS4 region has the nucleotide sequence set forth in SEQ ID NO:6, which is provided below:

45

50

55

TGAGCCCCCTTTTCTCTTAACCTGAAAGAAGGAAAAAAAAAATGGAACCCAAAATATTCTACATA
 GTTTCATGTACAGCCAGGGCTGGGCAGTCTCCTGTTATTTCTTTTAAAATAAATATATCAT
 TTAAATGCATAAATAAGCAAACCTGCTCGGGAATGGGAGGGAGAGTCTCTGGAGTCCACCCC
 5 TTCTCGGCCCTGGCTCTGCAGATAGTGCTATCAAAGCCCTGACAGAGCCCTGCCATTGCTGG
 GCCTTGGAGTGAGTCAGCCTAGTAGAGAGGCAGGGCAAGCCATCTCATAGCTGCTGAGTGGGA
 GAGAGAAAAGGGCTCATTGTCTATAAACTCAGGTCATGGCTATTCTTATTCTCACACTAAGAA
 AAAGAATGAGATGTCTACATATACCCTGCGTCCCCCTCTTGTGTACTGGGGCCCCCAAGAGCTC
 TCTAAAAGTGATGGCAAAGTCATTGCGCTAGATGCCATCCCATCTATTATAAACCTGCATTTG
 10 TCTCCACACACCAGTCATGGACAATAACCTCCTCCCAGGTCCACGTGCTTGTCTTTGTATAA
 TACTCAAGTAATTTTCGAAAAATGTATTCTTTCAATCTTGTCTGTTATTCTGTTTCAATGGC
 TTAGTAGAAAAAGTACATACTTGTGTTTTCCATAAATTGACAATAGACAATTTACATCAATGT
 CTATATGGGTCGTTGTGTTTGTCTGTGTTTGCAAAAACTCACAATAACTTTATATTGTTACTAC
 TCTAAGAAAGTTACAACATGGTGAATACAAGAGAAAGCTATTACAAGTCCAGAAAAATAAAGT
 TATCATCTTGAGGCCCTCAGCTTTCTAGGAATAATATCAATATTACAAAATTTAATCTAACAAAT
 15 TATGAACAGCAATGAGATAATATGTACAAAAGTACCCAGACCTATGTGGTAGAGCATCAAGGAA
 GCGCATTGCGGAGCAGTTTTTTGTTTGTGTTTGTGTTTGTATTCTGTTTCGTGAGGCAAGGTTTC
 ACTCTGCTGTCCAGGCTGGAGTGCAGTGGCAAGATCATGTCTCACTGCAGCCTTGAC [SEQ
 ID NO:6]

[0088] In one non-limiting example, the HS4 region has the nucleotide sequence set forth in SEQ ID NO:7, which is provided below:

TGAGCCCCCTTTTCTCTTAACCTGAAAGAAGGAAAAAAAAAATGGAACCCAAAATATTCTAC

ATAGTTTCCATGTACAGCCAGGGCTGGGCAGTCTCCTGTTATTTCTTTTAAAATAAATA
 TATCATTTAAATGCATAAATAAGCAAACCTGCTCGGGAATGGGAGGGAGAGTCTCTGGA
 GTCCACCCCTTCTCGGCCCTGGCTCTGCAGATAGTGCTATCAAAGCCCTGACAGAGCCCT
 30 GCCCATTTGCTGGGCCCTGGAGTGAGTCAGCCTAGTAGAGAGGCAGGGCAAGCCATCTCAT
 AGCTGCTGAGTGGGAGAGAGAGAAAAGGGCTCATTGTCTATAAACTCAGGTCATGGCTATTCT
 TTATTCTCACACTAAGAAAAAGAATGAGATGTCTACATATACCCTGCGTCCCCCTCTTGTG
 TACTGGGGCCCCCAAGAGCTCTCTAAAAGTGATGGCAAAGTCATTGCGCTAGATGCCATC
 CCATCTATTATAAACCTGCATTTGTCTCCACACACCAGTCATGGACAATAACCTCCTCC
 CAGGTCCACGTGCTTGTCTTTGTATAATACTCAAGTAATTTTCGAAAAATGTATTCTTTCA
 35 ATCTTGTCTGTTATTCTGTTTCAATGGCTTAGTAGAAAAAGTACATACTTGTGTTTCCC
 ATAAATTGACAATAGACAATTTACATCAATGTCTATATGGGTCGTTGTGTTTGTCTGTGT
 TTGCAAAAACTCACAATAACTTTATATTGTTACTACTCTAAGAAAGTTACAACATGGTGA
 ATACAAGAGAAAGCTATTACAAGTCCAGAAAATAAAGTTATCATCTTGAGGCCTCAGCT
 TTCTAGGATAAATATCAATATTACAAAATTAATCTAACAATTATGAACAGCAATGAGAT
 40 AATATGTACAAAGTACCCAGACCTATGTGGTAGAGCATCAAGGAAGCGCATTGCGGAGCA
 GTTTTTTGTGTTGTTTGTGTTTGTATTCTGTTTCGTGAGGCAAGGTTTCACTCTGCTGTCC
 AGGCTGGAGTGCAGTGGCAAGATCATGTCTCACTGCAGCCTTGACAC [SEQ ID NO:7]

[0089] In certain embodiments, the HS4 region has a length of less than about 1.0 kb, e.g., less than about 900 bp, less than about 700 bp, less than about 600 bp, or less than about 500 bp. In certain embodiments, the HS4 region has a length of less than about 500 bp. In certain embodiments, the HS4 region has a length of about 450 bp. In one non-limiting embodiment, the HS4 region has a length of about 446 bp. In one non-limiting example, the HS4 region has the nucleotide sequence set forth in SEQ ID NO:8, which is provided below:

TGAACCCAAAATATTCTACATAGTTTCCATGTACAGCCAGGGCTGGGCAGTCTCCTGTTAT
 TTCTTTTAAAATAAATATATCATTTAAATGCATAAATAAGCAAACCTGCTCGGGAATGGGAG
 GGAGAGTCTCTGGAGTCCACCCCTTCTCGGCCCTGGCTCTGCAGATAGTGCTATCAAAGCCCT
 GACAGAGCCCTGCCCATTTGCTGGGCCCTGGAGTGAGTCAGCCTAGTAGAGAGGCAGGGCAAGC
 55 CATCTCATAGCTGCTGAGTGGGAGAGAGAGAAAAGGGCTCATTGTCTATAAACTCAGGTCATGGC
 TATTCTTATTCTCACACTAAGAAAAAGAATGAGATGTCTACATATACCCTGCGTCCCCCTCTTG
 TGTACTGGGGTCCCCAAGAGCTCTCTAAAAGTGATGGCAAAGTCATTGCGCTAGATGCCATCC
 CATCT [SEQ ID NO:8]

[0090] In certain embodiments, the HS4 region has a length of about 280 bp (e.g., 283 bp). In certain embodiments, the HS4 region has a length of about 240 bp (e.g., 243 bp).

[0091] In certain non-limiting embodiments, the β -globin LCR region comprises a HS2 region having the nucleotide sequence set forth in SEQ ID NO:9, SEQ ID NO:20 or SEQ ID NO:21, a HS3 region having the nucleotide sequence set forth in SEQ ID NO:5, and a HS4 region having the nucleotide sequence set forth in SEQ ID NO:6, SEQ ID NO:7 or SEQ ID NO:8.

[0092] In one non-limiting embodiment, the β -globin LCR region comprises a HS2 region having the nucleotide sequence set forth in SEQ ID NO:9, a HS3 region having the nucleotide sequence set forth in SEQ ID NO:5, and a HS4 region having the nucleotide sequence set forth in SEQ ID NO:7, as shown in Figure 1.

[0093] In another non-limiting embodiment, the β -globin LCR region further comprises a HS1 region, *i.e.*, a β -globin LCR region comprising a HS1 region, a HS2 region, a HS3 region, and a HS4 region. In certain embodiments, the HS1 region, HS2 region, HS3 region and HS4 region within the β -globin LCR region are contiguous. In one non-limiting embodiment, the β -globin LCR region consisting essentially of a HS1 region, a HS2 region, a HS3 region and a HS4 region. In another embodiment, the β -globin LCR region comprises two introduced GATA-1 binding sites at the junction between the HS3 region and the HS4 region.

[0094] The length and the sequence of the HS1 region can vary. In certain embodiments, the HS1 region is from about 300 bp to about 1500 bp in length, e.g., from about 300 bp to about 1100 bp in length. In certain embodiments, the HS1 region has a length of about 1.0 kb or more, e.g., about 1.1 kb, about 1.2 kb, about 1.3 kb, about 1.4 kb, or about 1.5kb. In certain embodiments, the HS1 region has a length of about 1.1 kb. In one non-limiting example, the HS1 region has a length of 1074 bp. In one non-limiting example, the HS1 region has the nucleotide sequence set forth in SEQ ID NO:2, which is provided below:

AAGTAAACTTCCACAACCGCAAGCTTATTGAGGCTAAGGCATCTGTGAAGGAAAGAACATCTCCTCTA
AACCACATATGCTGCTAGAGCCTCTTTTCTGTACTCAAGCCTCATTCAGACACTAGTGTACCAGTCTCC
TCATATACCTATTGTATTTTCTTCTTCTTGCTGGTTTAGTCATGTTTTCTGGGAGCTTAGGGGCTTATT
TTATTTTGTTTTGTTCCTAATCAACAGAGATGGGCAAACCCATTATTTTTTCTTTAGACTTGGGATG
GTGATAGCTGGGCAGCGTCAGAACTGTGTGTGGATATAGATAAGAGCTCGGACTATGCTGAGCTGTGA
TGAGGGAGGGACCTAGCCAAAGGCAGTGAGAGTCAGAATGCTCCTGCTATTGCCTTCTCAGTCCCCAG
CTTGGTTTCTACACAAGTAGATACATAGAAAAGGCTATAGGTTAGTGTGTTGAGAGTCCTGCATGAGTTA
GTTGCTCAGAAATGCCCCGATAAATATGTTATGTGTGTTTATGTATATATATGTTTTATATATATATATG
TGTGTGTGTGTGTGTGTGTGTGTTGTGTTTACAAATATGTGATTATCATCAAACGTGAGGGCTAAAGT
GACCAGATAACTTGCAGGTCTTAGGATACCAGGAAAATAAATTACATTCCAAAATTTAACTGAGACTT
TAAAAAAAAAAAAAAAAAAAAAAAAAAAAACCAGTGATCCATGGACACAGGGAGGGGAACATCACACACT
GGGGCCTGTTGGGGGTGGGGGGCTAGGGGAAGGATAGCATTAGGAGAAATACCTAATGTAGATGACGGG
TTGATGGGTGCAGCAAACCACCATGGCACATGTACCCAGAACTTAAAGCATATTAAAAAACAGTGAT
CATAAAAGAAGCTCAAATTTAACTATAAGAGACGGAATGGCTCCCACAATTCTTAACTATAATCTTACA
GAATATTCTCATTTGAATAGAAGTATGCTTATCATTAGAGATTTGGACAGCCAGGAAAGCACAGAAAAAA
AAAAAAGGAGCTCTGTTGCCTTATAGCCTAGAGGTGTTT [SEQ ID NO:2]

[0095] In certain embodiments, the HS1 region has a length of less than about 1.0 kb, e.g., from about 400 bp to about 700 bp, from about 400 bp to about 500 bp, from about 500 bp to about 600 bp, from about 600 bp to about 700 bp, from about 700 bp to about 800 bp, from about 800 bp to about 900 bp, or from about 900 bp to about 1.0 kb. In certain embodiments, the HS1 region has a length of less than about 700 bp. In certain embodiments, the HS1 region has a length of about 600 bp. In one non-limiting embodiment, the HS1 region has a length of 602 bp. In one non-limiting example, the HS1 region has the nucleotide sequence set forth in SEQ ID NO:3, which is provided below:

GGCATCTGTGAAGGAAAGAAACATCTCCTCTAAACCACTATGCTGCTAGAGCCTCTTTTCTGTACTCAA
GCCTCATTGAGACACTAGTGTACCAGTCTCCTCATATACCTATTGTATTTTCTTCTTCTTGCTGGTTT
AGTCATGTTTTCTGGGAGCTTAGGGGCTTATTTTATTTTGTTTTGTTTTCTAATCAACAGAGATGGGCA
AACCCATTATTTTTTTTCTTTAGACTTGGGATGGTGATAGCTGGGCAGCGTCAGAAACTGTGTGTGGATA
TAGATAAGAGCTCGGACTATGCTGAGCTGTGATGAGGGAGGGACCTAGCCAAAGGCAGTGAGAGTCAGA
ATGCTCCTGCTATTGCCTTCTCAGTCCCCACGCTTGGTTTTCTACACAAGTAGATACATAGAAAAGGCTA
TAGGTTAGTGTTTGAGAGTCCTGCATGAGTTAGTTGCTCAGAAATGCCCGATAAATATGTTATGTGTGT
TTATGTATATATATGTTTTATATATATATATGTGTGTGTGTGTGTGTGTGTGTGTTGTGTTTACAAATA
TGTGATTATCATCAAAACGTGAGGGCTAAAGTGACCAGATAACTTGCAGG [SEQ ID NO:3]

[0096] In certain embodiments, the HS1 region has a length of less than about 500 bp. In certain embodiments, the HS1 region has a length of about 490 bp. In one non-limiting embodiment, the HS1 region has a length of 489 bp. In one non-limiting example, the HS1 region has the nucleotide sequence set forth in SEQ ID NO:4, which is provided below:

```

5      GGCATCTGTGAAGGAAAGAAACATCTCCTCTAAACCACTATGCTGCTAGAGCCTCTTTTCTGTACTCAA
      GCCTCATTCAGACACTAGTGTCAACAGTCTCCTCATATACCTATTGTATTTTCTTCTTCTTGCTGGTTT
      AGTCATGTTTTCTGGGAGCTTAGGGGCTTATTTTATTTTGTGTTTTCTAATCAACAGAGATGGGCA
      AACCCATTATTTTTTCTTTAGACTTGGGATGGTGATAGCTGGGCAGCGTCAGAACTGTGTGTGGATA
10     TAGATAAGAGCTCGGACTATGCTGAGCTGTGATGAGGGAGGGACCTAGCCAAAGGCAGTGAGAGTCAGA
      ATGCTCCTGCTATTGCCCTTCTCAGTCCCCACGCTTGGTTTCTACACAAGTAGATACATAGAAAAGGCTA
      TAGGTTAGTGTTTGTAGAGTCCCTGCATGAGTTAGTTGCTCAGAAATGCCCCGATAAATATGTTATGTGTGT
      TTATGT [SEQ ID NO:4]

```

[0097] Recent studies have shown that HS2 is not erythroid-specific, but is expressed in other cell lines and lineages (See Example 3 and Figure 7) and is also present in undifferentiated human embryonic stem cells (Chang et al., Stem cell reviews (2013);9:397-407). Due to the non-erythroid activity of HS2, HS2-containing globin vectors may pose a risk for their safe use in clinical treatment, e.g., for treating thalassemia and sickle cell patients. In certain embodiments, the β -globin LCR region does not comprise a HS2 region. In certain embodiments, the β -globin LCR region does not comprise a core sequence of HS2. A core sequence of HS2 provides position independent, high level expression. In addition, a core sequence of HS2 sustains the enhancer activity of HS2. For example, the core sequence of HS2 enhances the transcription of a globin gene (e.g., human β -globin gene). Additionally, a core sequence of HS2 comprises one or more binding sites or binding motifs for ubiquitous as well as tissue-specific (e.g., erythroid-specific) proteins (e.g., transcription factors), including, but not limited to, members of API family of proteins (e.g., NF-E2), GATA-1 (also known as "NF-E1" or "NFE1"), Kruppel-like Zn finger proteins (e.g., ubiquitous proteins Sp1 and YY1, and erythroid-restricted factor erythroid Kruppel-like factor (EKLF)), and basic helix-loop-helix (bHLH) proteins (E boxes) (e.g., USF and TAL1). AP1 binding sites are required for enhancement and induction (Moi and Kan (1990); Ney et al., (1990); Talbot and Grosveld (1991)). Furthermore, binding of NF-E2 can cause disruption of *in vitro* reconstituted chromatin at HS2 (Armstrong and Emerson (1996)). Mutations in the GATA-1 binding sites can cause a reduction in enhancer activity of HS2 in transgenic mice (Caterina et al., (1994)). Although both API (e.g., AP1/NF-E2) and GATA1 binding sites are important for core function, mice lacking these factors do not show impaired globin gene expression (Weiss et al., 1994).

[0098] In certain embodiments, the β -globin LCR region does not comprise the full length of a core sequence of HS2. In certain embodiments, the core sequence of a HS2 region is a core sequence of human HS2. In one non-limiting embodiment, the core sequence of human HS2 comprises a tandem pair of binding sites for members of API family of proteins (e.g., NF-E2) (referred to as "AP1/NF-E2" binding sites) (e.g., GCTGAGTCA, and GATGAGTCA), one binding site for Kruppel-like Zn finger proteins (e.g., AGGGTGTGT), one GATA-1 binding site (e.g., CTATCT), and three E boxes (CANNTG, e.g., CAGATG, and CACCTG). In one non-limiting embodiment, the β -globin LCR region does not comprise the full length of a 388bp core sequence of human HS2, which has the nucleotide sequence set forth in SEQ ID NO:20 provided below:

```

40     TAAGCTTCAGTTTTTCTTAGTTCCTGTTACATTTCTGTGTGTCTCCATTAGTGACCTCCCATAGTCCA
      AGCATGAGCAGTTCTGGCCAGGCCCTGTGCGGGTCAGTGCCCCACCCCGCCTTCTGGTTCTGTGTAA
      CCTTCTAAGCAAACCTTCTGGCTCAAGCACAGCAATGCTGAGTCATGATGAGTCATGCTGAGGCTTAGG
45     GTGTGTGCCCAGATGTTCTCAGCCTAGAGTGATGACTCCTATCTGGGTCCCCAGCAGGATGCTTACAGG
      GCAGATGGCAAAAAAAGGAGAAGCTGACCACCTGACTAAACTCCACCTCAAACGGCATCATAAAGAA
      AATGGATGCCTGAGACAGAATGTGACATATTCTAGAATATATT [SEQ ID NO:20]

```

[0099] The nucleotide sequence set forth in SEQ ID NO:20 corresponds to nucleotides position 16671 to position 17058 of SEQ ID NO:19 (GenBank Access No.: NG_000007.3). In SEQ ID NO:20, one AP1/NF-E2 binding site having the nucleotide sequence of GCTGAGTCA is located at position 175 to position 183, one AP1/NF-E2 binding site having the nucleotide sequence of GATGAGTCA is located at position 185 to position 193, one binding site for Kruppel-like Zn finger proteins having the nucleotide sequence of AGGGTGTGT is located as position 205 to position 213, two E boxes, each of which have the nucleotide sequence of CAGATG, is located at position 217 to position 222, and position 278 to position 283, one GATA-1 binding site having the nucleotide sequence of CTATCT is located at position 246 to position 251, one E box having the nucleotide sequence of CACCTG is located at position 306 to position 311.

[0100] In one non-limiting embodiment, the β -globin LCR region does not comprise the full length of a 387bp core sequence of human HS2, which has the nucleotide sequence set forth in SEQ ID NO:21 provided below:

TAAGCTTCAGTTTTTCCTTAGTTCCTGTTACATTTCTGTGTGTCTCCATTAGTGACCTCCCATAGTCCA
 AGCATGAGCAGTTCTGGCCAGGCCCTGTCGGGGTCAGTGCCCCACCCCCGCCTTCTGGTTCTGTGTAA
 CCTTCTAAGCAAACCTTCTGGCTCAAGCACAGCAATGCTGAGTCATGATGAGTCATGCTGAGGCTAGGG
 TGTGTGCCCAGATGTTCTCAGCCTAGAGTGATGACTCCTATCTGGGTCCCCAGCAGGATGCTTACAGGG
 CAGATGGCAAAAAAAGGAGAAGCTGACCACCTGACTAAAACCTCCACCTCAAACGGGCATCATAAAGAAA
 ATGGATGCCTGAGACAGAATGTGACATATTCTAGAATATATT [SEQ ID NO:21]

[0101] In SEQ ID NO:21, one AP1/NF-E2 binding site having the nucleotide sequence of GCTGAGTCA is located at position 175 to position 183, one AP1/NF-E2 binding site having the nucleotide sequence of GATGAGTCA is located at position 185 to position 193, one binding site for Kruppel-like Zn finger proteins having the nucleotide sequence of AGGGTGTGT is located as position 204 to position 212, two E boxes, each of which have the nucleotide sequence of CAGATG, is located at position 216 to position 221, and position 277 to position 282, one GATA-1 binding site having the nucleotide sequence of CTATCT is located at position 245 to position 250, one E box having the nucleotide sequence of CACCTG is located at position 305 to position 310.

[0102] In certain embodiments, the β -globin LCR region does not comprise a HS2 region that comprises a core sequence of HS2. A HS2 region that comprises a core sequence of HS2 can vary in length and sequence. In non-limiting examples, a HS2 region that comprises a core sequence of HS2 is from about 400 bp to about 1000 bp, e.g., from about 400 bp to about 500 bp, from about 500 bp to about 600 bp, from about 600 bp to about 700 bp, from about 700 bp to about 800 bp, from about 800 bp to about 900 bp, or from about 900 bp to about 1000 bp, in length. In one non-limiting embodiment, the β -globin LCR region does not comprise a 840 bp HS2 region (e.g., the HS2 region comprised in the globin vector TNS9 disclosed in U.S. 7,541,179). In one non-limiting embodiment, the β -globin LCR region does not comprise a 860 bp HS2 region. In one non-limiting embodiment, the β -globin LCR region does not comprise an about 650 bp HS2 region. In one non-limiting example, the β -globin LCR region does not comprise a 646 bp HS2 region (e.g., the HS2 region comprised in the globin vector LentiGlobin™, also known as " β^{87} "). In one non-limiting embodiment, the β -globin LCR region does not comprise an about 420 bp HS2 region. In one non-limiting example, the β -globin LCR region does not comprise a 423 bp HS2 region (e.g., the HS2 region comprised in the globin vector disclosed in Sadelain et al., Proc. Nat'l Acad. Sci. (USA) (1995);92:6728-6732).

[0103] In certain embodiments, the β -globin LCR region does not comprise a HS2 region that sustains the enhancer activity of HS2. In certain embodiments, the β -globin LCR region does not comprise a HS2 region that is capable of enhancing the transcription of a globin gene (e.g., human β -globin gene). In non-limiting examples, the β -globin LCR region does not comprise a HS2 region whose ability to enhance the transcription of a globin gene (e.g., human β -globin gene) is no less than 60%, no less than 70%, no less than 80%, no less than 90%, or no less than 95% in comparison to a native HS2 region.

[0104] In certain embodiments, the β -globin LCR region does not comprise a HS2 region that comprises one, two, three, four, five, six or seven of the following binding sites: two (a tandem pair of) AP1/NF-E2 binding sites (e.g., GCTGAGTCA, and GATGAGTCA), one binding site for Kruppel-like Zn finger proteins (e.g., AGGGTGTGT), one GATA-1 binding site (e.g., CTATCT), and three E boxes (CANNTG, e.g., CAGATG, and CACCTG). In certain embodiments, the β -globin LCR region does not comprise a HS2 region that comprises six of the above-described binding sites. For example, in certain embodiments, the β -globin LCR region does not comprise a HS2 region that comprises two AP1/NF-E2 binding sites, one binding site for Kruppel-like Zn finger proteins, one GATA-1 binding site, and two not three E boxes. In certain embodiments, the β -globin LCR region does not comprise a HS2 region that comprises one not two AP1/NF-E2 binding site, one binding site for Kruppel-like Zn finger proteins, one GATA-1 binding site, and three E boxes. In certain embodiments, the β -globin LCR region does not comprise a HS2 region that comprises two AP1/NF-E2 binding sites, one GATA-1 binding site, and three E boxes and does not comprise one binding site for Kruppel-like Zn finger proteins. In certain embodiments, the β -globin LCR region does not comprise a HS2 region that comprises two AP1/NF-E2 binding sites, one binding site for Kruppel-like Zn finger proteins, and three E boxes, and does not comprise one GATA-1 binding site.

[0105] In certain embodiments, the β -globin LCR region comprises a HS1 region, a HS3 region, and a HS4 region, and does not comprise a HS2 region. In certain embodiments, the HS1 region, HS3 region and HS4 region within the β -globin LCR region are contiguous. In one non-limiting embodiment, the β -globin LCR region consisting essentially of a HS1 region, a HS3 region and a HS4 region. In another embodiment, the β -globin LCR region comprises two introduced GATA-1 binding sites at the junction between the HS3 region and the HS4 region. The HS3 region can lie between the

HS1 region and the HS4 region.

[0106] In certain non-limiting embodiments, the β -globin LCR region comprises a HS1 region having the nucleotide sequence set forth in SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:22 or SEQ ID NO:23, a HS3 region having the nucleotide sequence set forth in SEQ ID NO:5, and a HS4 region having the nucleotide sequence set forth in SEQ ID NO:6, SEQ ID NO:7 or SEQ ID NO:8, and the β -globin LCR region does not comprise a HS2 region.

[0107] In one non-limiting embodiment, the β -globin LCR region comprises a HS1 region having the nucleotide sequence set forth in SEQ ID NO:2, a HS3 region having the nucleotide sequence set forth in SEQ ID NO:5, and a HS4 region having the nucleotide sequence set forth in SEQ ID NO:6, and the β -globin LCR region does not comprise a HS2 region, as shown in Figure 2.

[0108] In one non-limiting embodiment, the β -globin LCR region comprises a HS1 region having the nucleotide sequence set forth in SEQ ID NO:3, a HS3 region having the nucleotide sequence set forth in SEQ ID NO:5, and a HS4 region having the nucleotide sequence set forth in SEQ ID NO:8, and the β -globin LCR region does not comprise a HS2 region, as shown in Figure 3.

[0109] In one non-limiting embodiment, the β -globin LCR region comprises a HS1 region having the nucleotide sequence set forth in SEQ ID NO:4, a HS3 region having the nucleotide sequence set forth in SEQ ID NO:5, and a HS4 region having the nucleotide sequence set forth in SEQ ID NO:8, and the β -globin LCR region does not comprise a HS2 region.

[0110] In certain embodiments, the β -globin LCR region does not comprise a HS1 region or a HS2 region. In certain embodiments, the β -globin LCR region does not comprise a core sequence of HS1. A core sequence of HS1 sustains the activity of HS1, e.g., enhancer activity, or functioning as a facilitator or regulatory element to tether the enhancer activity of other HS regions, e.g., HS2-4. In addition, a core sequence of HS1 comprises one or more binding sites or binding motifs for ubiquitous as well as tissue-specific (e.g., erythroid-specific) proteins (e.g., transcription factors), including, but not limited to, GATA-1, and Kruppel-like Zn finger proteins (e.g., erythroid-restricted factor EKLF).

[0111] In certain embodiments, the β -globin LCR region does not comprise the full length of a core sequence of HS1. In certain embodiments, the core sequence of a HS1 region is a core sequence of human HS1. In one non-limiting embodiment, the core sequence of human HS1 comprises two GATA-1 binding sites (e.g., TTATCT, and CTATCA), and one binding site for EKLF (e.g., CCACACACA). In certain embodiments, the β -globin LCR region does not comprise the full length of a 286bp core sequence of human HS1. In one non-limiting embodiment, the 286bp core sequence of human HS1 has the nucleotide sequence set forth in SEQ ID NO:22 provided below:

```
CTGAGCAACTAACTCATGCAGGACTCTCAAACACTAACCTATAGCCTTTTCTATGTATCTACTTGTGTGA
GAAACCAAGCGTGAGGACTGAGAAGGCAATAGCAGGAGCATTCTGACTCTCACTGCCTTTGGCTAGGTC
CCTCCCTCATCACAGCTCAGCATAGTCCGAGCTCTTATCTATATCCACACACAGTTTCTGACGCTGCCC
AGCTATCACCATCCCAAGTCTAAAGAAAAAATAATGGGTTTGCCCATCTCTGTTGATTAGAAAACAAA
ACAAAATAAA [SEQ ID NO:22]
```

In SEQ ID NO:22, one GATA-1 binding site having the nucleotide sequence of TTATCT is located at position 173 to position 178, one GATA-1 binding site having the nucleotide sequence of CTATCA located at position 210 to position 215, and one binding site for EKLF having the nucleotide sequence of CCACACACA is located at position 183 to position 191.

[0112] In another non-limiting embodiment, the 286bp core sequence of human HS1 has the nucleotide sequence set forth in SEQ ID NO:23 provided below:

```
CTGAGCAACTAATCATGCAGGACTCTCAAACACTAACCTATAGCCTTTTCTATGTATCTACTTGTGTAG
AAACCAAGCGTGAGGACTGAGAAGGCAATAGCAGGAGCATTCTGACTCTCACTGCCTTTAGCTAGGCC
CTCCCTCATCACAGCTCAGCATAGTCCGAGCTCTTATCTATATCCACACACAGTTTCTGACGCTGCCC
AGCTATCACCATCCCAAGTCTAAAGAAAAAATAATGGGTTTGCCCATCTCTGTTGATTAGAAAACAAA
ACAAAATAAA [SEQ ID NO:23]
```

[0113] The nucleotide sequence set forth in SEQ ID NO:23 corresponds to nucleotides position 21481 to position 21766 of SEQ ID NO:19 (GenBank Access No.: NG_000007.3). In SEQ ID NO:23, one GATA-1 binding site having the nucleotide sequence of TTATCT is located at position 173 to position 178, one GATA-1 binding site having the nucleotide

sequence of CTATCA located at position 210 to position 215, and one binding site for EKLF having the nucleotide sequence of CCACACACA is located at position 183 to position 191.

[0114] In certain embodiments, the β -globin LCR region does not comprise a HS1 region that comprises a core sequence of HS1. A HS1 region that comprises a core sequence of HS1 can vary in length and sequence. In non-limiting examples, a HS1 region that comprises a core sequence of HS1 is from about 300 bp to about 1200 bp, e.g., from about 300 bp to about 400 bp, from about 400 bp to about 500 bp, from about 500 bp to about 600 bp, from about 600 bp to about 700 bp, from about 700 bp to about 800 bp, from about 800 bp to about 900 bp, from about 900 bp to about 1000 bp, from about 1000 bp to about 1100 bp, or from about 1100 bp to about 1200 bp, in length. In one non-limiting embodiment, the β -globin LCR region does not comprise an about 1.0 kb bp HS1 region. In one non-limiting embodiment, the β -globin LCR region does not comprise an about 1.1 kb HS1 region.

[0115] In certain embodiments, the β -globin LCR region does not comprise a HS1 region that sustains the activity of HS1, e.g., enhancer activity, or functioning as a facilitator or regulatory element to tether the enhancer activity of other HS regions, e.g., HS2-4. In certain embodiments, the β -globin LCR region does not comprise a HS1 region that is capable of enhancing the transcription of a globin gene (e.g., human β -globin gene). In non-limiting examples, the β -globin LCR region does not comprise a HS1 region whose ability to enhance the transcription of a globin gene (e.g., human β -globin gene) is no less than 60%, no less than 70%, no less than 80%, no less than 90%, or no less than 95% in comparison to a native HS1 region. In non-limiting examples, the β -globin LCR region does not comprise a HS1 region whose ability to tether the enhancer activity of one or more of HS2-HS4 is no less than 60%, no less than 70%, no less than 80%, no less than 90%, or no less than 95% in comparison to a native HS1 region.

[0116] In certain embodiments, the β -globin LCR region does not comprise a HS1 region that comprises one, two, or three of the following binding sites: two GATA-1 binding sites (e.g., TTATCT, and CTATCA), and one binding site for EKLF (e.g., CCACACACA). In certain embodiments, the β -globin LCR region does not comprise a HS1 region that comprises two of the above-described binding sites. For example, in certain embodiments, the β -globin LCR region does not comprise a HS1 region that comprises two GATA-1 binding sites and does not comprise one binding site for EKLF. In certain embodiments, the β -globin LCR region does not comprise a HS1 region that comprises one not two AP1/NF-E2 binding site and one binding site for EKLF.

[0117] In certain embodiments, the β -globin LCR region comprises a HS3 region and a HS4 region, and the β -globin LCR region does not comprise a HS1 region or a HS2 region. In certain embodiments, the HS3 region and HS4 region within the β -globin LCR region are contiguous. In one non-limiting embodiment, the β -globin LCR region consisting essentially of a HS3 region and a HS4 region. In another embodiment, the β -globin LCR region comprises two introduced GATA-1 binding sites at the junction between the HS3 region and the HS4 region. The HS3 region can lie between the globin gene or functional portion thereof and the HS4 region.

[0118] In certain embodiments, the β -globin LCR region comprises a HS3 region having the nucleotide sequence set forth in SEQ ID NO:5 and a HS4 region having the nucleotide sequence set forth in SEQ ID NO:6, SEQ ID NO:7 or SEQ ID NO:8, and the β -globin LCR region does not comprise a HS1 region or a HS2 region.

[0119] In one non-limiting embodiment, the β -globin LCR region comprises a HS3 region having the nucleotide sequence set forth in SEQ ID NO:5 and a HS4 region having the nucleotide sequence set forth in SEQ ID NO:6, and the β -globin LCR region does not comprise a HS1 region or a HS2 region, as shown in Figure 4.

Globin gene

[0120] In accordance with the presently disclosed subject matter, the expression cassette comprises a globin gene or a functional portion thereof. The globin gene can be a β -globin gene, a γ -globin gene, or a δ -globin gene. In certain embodiments, the expression cassette comprises a human β -globin gene. In accordance with the presently disclosed subject matter, the human β -globin gene can be a wild-type human β -globin gene, a deleted human β -globin gene comprising one or more deletions of intron sequences, or a mutated human β -globin gene encoding at least one anti-sickling amino acid residue. In one non-limiting embodiment, a presently disclosed expression cassette comprises a wild-type human β -globin gene. In another embodiment, the a presently disclosed expression cassette comprises a human β^A -globin gene encoding a threonine to glutamine mutation at codon 87 (β^A -T87Q). The glutamine residue at position 87 in the gamma-globin chain augments the anti-sickling activity of the gamma chain relative to the beta chain, while preserving adult oxygen-binding characteristics of the beta chain (Nagel et al., Proc. Natl. Acad. Sci. U.S.A. (1979);76:670-672). In certain embodiments, a functional portion of a globin gene has at least 80%, at least 90%, at least 95%, or at least 99% identity to a corresponding wild-type reference polynucleotide sequence.

Promoters and Enhancers

[0121] In accordance with the presently disclosed subject matter, the expression cassette can further comprise a β -globin promoter. In certain embodiments, the β -globin promoter is positioned between the globin gene or functional

portion thereof and the β -globin LCR region. The length and the sequence of the β -globin promoter can vary. In certain embodiments, the β -globin promoter is from about 100 bp to about 1600 bp in length, e.g., from about 200 bp to about 700 bp, from about 100 bp to about 200 bp, from about 200 bp to about 300 bp, from about 300 bp to about 400 bp, from about 400 bp to about 500 bp, from about 500 bp to about 600 bp, from about 600 bp to about 700 bp, from about 700 bp to about 800 bp, from about 800 bp to about 900 bp, from about 900 bp to about 1000 bp, from about 1000 bp to about 1100 bp, from about 1100 bp to about 1200 bp, from about 1200 bp to about 1300 bp, from about 1300 bp to about 1400 bp, from about 1400 bp to about 1500 bp, or from about 1500 bp to about 1600 bp in length. In certain embodiments, the β -globin promoter a human β -globin promoter that is about 130bp, about 613bp, about 265bp, or about 1555bp, in length. In one embodiment, the β -globin promoter is a human β -globin promoter that is about 613bp in length. In one non-limiting example, the human β -globin promoter has the nucleotide sequence set forth in SEQ ID NO:10, which is provided below:

```

AAGCAATAGATGGCTCTGCCCTGACTTTTATGCCCAGCCCTGGCTCCTGCCCTCCCTGCTCCT
GGGAGTAGATTGGCCAACCCCTAGGGTGTGGCTCCACAGGGTGAGGTCTAAGTGATGACAGCCG
TACCTGTCCTTGGCTCTTCTGGCACTGGCTTAGGAGTTGGACTTCAAACCCCTCAGCCCTCCCT
CTAAGATATATCTCTTGGCCCCATACCATCAGTACAAATTGCTACTAAAAACATCCTCCTTTG
CAAGTGTATTTACGTAATATTTGGAATCACAGCTTGGTAAGCATATTGAAGATCGTTTTCCCA
ATTTTCTTATTACACAAATAAGAAATTGATGCACTAAAAGTGGAAGAGTTTTGTCTACCATAA
TTCAGCTTTGGGATATGTAGATGGATCTCTTCTGCGTCTCCAGAATATGCAAAATACTTACA
GGACAGAATGGATGAAAACCTCTACCTCAGTTCTAAGCATATCTTCTCCTTATTTGGATTAAAA
CCTTCTGGTAAGAAAAGAAAAAATATATATATATATGTGTATATATACACACATACATATA
CATATATATGCATTCATTTGTTGTTGTTTTTCTTAATTTGCTCATG [SEQ ID NO:10]

```

[0122] In one embodiment, the β -globin promoter is a human β -globin promoter that is about 265bp in length. In one non-limiting example, the human β -globin promoter has the nucleotide sequence set forth in SEQ ID NO:11.

```

AAGCAATAGATGGCTCTGCCCTGACTTTTATGCCCAGCCCTGGCTCCTGCCCTCCCTGCTCCT
GGGAGTAGATTGGCCAACCCCTAGGGTGTGGCTCCACAGGGTGAGGTCTAAGTGATGACAGCCG
TACCTGTCCTTGGCTCTTCTGGCACTGGCTTAGGAGTTGGACTTCAAACCCCTCAGCCCTCCCT
CTAAGATATATCTCTTGGCCCCATACCATCAGTACAAATTGCTACTAAAAACATCCTCCTTTG
CAAGTGTATTTAC [SEQ ID NO:11]

```

[0123] Additionally or alternatively, a presently disclosed expression cassette can further comprise a human β -globin 3' enhancer. In certain embodiments, the human β -globin 3' enhancer is positioned in the upstream of the globin gene or functional portion thereof. In certain embodiments, the β -globin 3' enhancer is from about 500 bp to about 1000 bp in length, e.g., from about 500 bp to about 600 bp, from about 600 bp to about 700 bp, from about 700 bp to about 800 bp, or from about 800 bp to about 900 bp in length. In one embodiment, the human β -globin 3' enhancer is about 879 bp in length. In one example, the human β -globin 3' enhancer has the nucleotide sequence set forth in SEQ ID NO:12.

```

TAGGTATTGAATAAGAAAAATGAAGTTAAGGTGGTTGATGGTAACACTATGCTAATAAC
TGCAGAGCCAGAAGCACCATAAGGGACATGATAAGGGAGCCAGCAGACCTCTGATCTCTT
CCTGAATGCTAATCTTAAACATCCTGAGGAAGAATGGGACTTCCATTTGGGGTGGGCCTA
TGATAGGGTAATAAGACAGTAGTGAATATCAAGCTACAAAAAGCCCCCTTCAAATTCTT
CTCAGTCCTAACTTTTCATACTAAGCCCAGTCTTCCAAAGCAGACTGTGAAAGAGTGAT
AGTTCCGGGAGACTAGCACTGCAGATTCCGGGTCACGTGTGAGTGGGGGAGGCAGGGAAGA
AGGGCTCACAGGACAGTCAAACCATGCCCCCTGTTTTTCTTCTTCAAGTAGACCTCTAT
AAGACAACAGAGACAATAAGGCTGAGTGGCCAGGCGAGGAGAAACCATCTCGCCGTAAA
ACATGGAAGGAACACTTCAGGGGAAAGGTGGTATCTCTAAGCAAGAGAACTGAGTGGAGT
CAAGGCTGAGAGATGCAGGATAAGCAAATGGGTAGTGAAGAGACATTCATGAGGACAGCT
AAAACAATAAGTAATGTAAATACAGCATAGCAAACTTTAACCTCCAAATCAAGCCTCT
ACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTG
CATTAGCTGTTTGCAGCCTCACCTTCTTTTATGTTTAAATGCACTGACCTCCCACATTCCCT
TTTGTAGTAAATATTCAGAAATAATTTAAATACATCATG [SEQ ID NO:12]

```

[0124] Furthermore, a presently disclosed expression cassette can further comprise at least one erythroid-specific enhancer. The presently disclosed expression cassette allows for expression of a globin gene (*e.g.*, human β -globin gene) in erythroid-specific fashion. The erythroid-specific enhancer can enhance the expression of the globin gene in erythroid-specific fashion. For example, the erythroid-specific enhancer lack enhancer activity in non-erythroid tissues. In particularly, for the β -globin LCR region that lacks a HS2 region, which primarily functions as an expression enhancer, the addition of one or more erythroid-specific enhancers can compensate the enhancing activity of a HS2 region. Furthermore, the presently disclosed erythroid-specific enhancers do not decrease or reduce the titer of a vector comprising the expression cassette. The length of the erythroid-specific enhancer can vary, *e.g.*, from about 100bp to about 200bp, from about 100bp to about 120bp, from about 120bp to about 140bp, from about 140 bp to about 200 (*e.g.*, from about 140bp to about 150bp, from about 150bp to about 160bp, from about 160bp to about 170bp, from about 170bp to about 180bp, from about 180bp to about 190bp, or from about 190bp to about 200bp). In certain embodiments, the erythroid-specific enhancer has a length of from about 140bp to about 200bp. In one non-limiting embodiment, the erythroid-specific enhancer has a length of 152 bp, which has the nucleotide sequence set forth in SEQ ID NO:13, which is provided below:

TCTCCACGCCCTGGTCTCAGCTTGGGGAGTGGTCAGACCCCAATGGCGATAAACTCTGGCAA
CTTTATCTGTGcaCTGCAGGCTCAGCCCCAAcaGCTTTAGCTTTCACAAGCAGGCAGGGGAAG
GGAAACACATATCTCCAGATATGAGG [SEQ ID NO:13]

[0125] In one non-limiting embodiment, the erythroid-specific enhancer has a length of 157 bp, which has the nucleotide sequence set forth in SEQ ID NO:14, which is provided below:

CTAAACCCCTCCCCCACCCTAGCCCCAAGCTTCATCTTAGCTCCACTCCTGACCCTATCCAGC
TAAAGGTCCCCACCCAGCTCCTGCCTATCTAGTCATTGCATATGGCAAGACTTGAAAGTCCTA
TCTCAAAGCAGCAGAATTATCAGCTACGACT [SEQ ID NO:14]

[0126] In one non-limiting embodiment, the erythroid-specific enhancer has a length of 141 bp, which has the nucleotide sequence set forth in SEQ ID NO:15, which is provided below:

CCATCCCCCAGCACTCCCTGCCCCCAGGCCAGACTTGACCAACTCCCAGCTccGCCTGGGA
CTTCCAGATATGGGGCCCCACCCTTGCCAGGCCTTGGGGACGCTGAAGATATTGACTATCTGCG
TGCCggAAAAGGGTG [SEQ ID NO:15]

[0127] In one non-limiting embodiment, the erythroid-specific enhancer has a length of 171 bp, which has the nucleotide sequence set forth in SEQ ID NO:16, which is provided below:

AAAGGCTGGGGGTGGGAGTAGCGGATTTGAAGCACTTGTGGCCTACAGAGGTGTGGCAAGCA
GAGCACCTCAGAACTCAGGCGTACTGCCCCGCCGCCGAGCCCTGCGAGGGCCGATAGCGAGGG
TGTGGCCCTTATCTGCACCCAGCAGAGCGCCGGCGGGGTACGGTC [SEQ ID NO:16]

[0128] In one non-limiting embodiment, the erythroid-specific enhancer has a length of 195 bp, which has the nucleotide sequence set forth in SEQ ID NO:17, which is provided below:

CAGTTGCCTCAGCTGAGTATGTCTTCTAAAGATAATGTCGATTGTGTATGGCTGATGGGATTC
TAGGACCAAGCAAGAGGTTTTTTTTTTTCCCCACATACTTAACGTTTCTATATTTCTATTTG
AATTGACTGGACAGTTCCATTGAATTATTTCTCTCTCTCTCTCTGACACATTTTATC
TTGCCA [SEQ ID NO:17]

[0129] Erythroid-specific enhancers can be identified and determined by any suitable methods known in the art. The erythroid-specific enhancers can be positioned at the 3' LTR (downstream) or the 5' LTR (downstream) of the β -globin LCR region. In one embodiment, the at least one erythroid-specific enhancer is positioned in the 5' LTR of the β -globin LCR region, *e.g.*, the upstream of the HS3 region. The expression cassette can comprise one, two, three, four, or five erythroid-specific enhancers. In one embodiment, the expression cassette comprises one erythroid-specific enhancer.

In another embodiment, the expression cassette comprises two erythroid-specific enhancers. In yet another embodiment, the expression cassette comprises three erythroid-specific enhancers. In certain embodiments, the expression cassette comprises four erythroid-specific enhancers. In a non-limiting embodiment, the expression cassette comprises five erythroid-specific enhancers.

Insulators

[0130] In accordance with the presently disclosed subject matter, the expression cassette comprises at least one of the above-described insulators. In certain embodiments, a presently disclosed expression cassette comprises at least one insulator comprising the CTCF binding site sequence set forth in SEQ ID NO:18, for example, but not limited to, an insulator comprising SEQ ID NO: 24 or SEQ ID NO:25, such as an insulator having the nucleotide sequence set forth in SEQ ID NO:1 (*i.e.*, insulator A1). In various non-limiting embodiments, the insulator can be incorporated or inserted into one or both LTRs or elsewhere in the region of a presently disclosed expression cassette that integrates into the cellular genome. In one embodiment, the insulator is positioned at the 3' end of the expression cassette. In one embodiment, the insulator is positioned at the 5' end of the expression cassette. In one embodiment, the expression cassette comprises two of the insulator having the nucleotide sequence set forth in SEQ ID NO:1, where one insulator is positioned at the 3' end and the other insulator is positioned at the 5' end of the expression cassette.

[0131] The presently disclosed insulators possess powerful enhancer blocking activity. In certain embodiments, the insulators possess barrier activity in addition to enhancer blocking activity. The presently disclosed insulators substantially decrease the risks of insertional mutagenesis and genotoxicity associated with viral vectors. Furthermore, when a presently disclosed insulator is incorporated into a vector, the insulator does not adversely effect vector titers of the vector. In certain embodiments, the insulators (*e.g.*, insulator A1) increase the *in vivo* expression of the globin gene or functional portion thereof.

[0132] For the purpose of illustration and not limitation, Figures 1-4 show recombinant vectors comprising exemplary expression cassettes in accordance with certain embodiments of the presently disclosed subject matter. Figure 1 shows a recombinant vector comprising a presently disclosed expression cassette that comprises a human β^{A-T87Q} globin gene, which is operably linked to a β -globin LCR region that comprises a 860bp HS2 region (*e.g.*, one having the nucleotide sequence set forth in SEQ ID NO:9), a 1301bp HS3 region (*e.g.*, one having the nucleotide sequence set forth in SEQ ID NO:5), and a 1065bp HS4 region (*e.g.*, one having the nucleotide sequence set forth in SEQ ID NO:7).

[0133] Figure 2 shows one exemplary recombinant vector comprising an expression cassette in accordance with one embodiment of the presently disclosed subject matter. Figure 2 shows a recombinant vector comprising a presently disclosed expression cassette that comprises a human β^{A-T87Q} globin gene, which is operably linked to a β -globin LCR region that comprises a 1.1 kb HS1 region (*e.g.*, one having the nucleotide sequence set forth in SEQ ID NO:2), a 1301bp HS3 region (*e.g.*, one having the nucleotide sequence set forth in SEQ ID NO:5), and a 1065 bp HS4 region (*e.g.*, one having the nucleotide sequence set forth in SEQ ID NO:6).

[0134] Figure 3 shows one exemplary recombinant vector comprising an expression cassette in accordance with one embodiment of the presently disclosed subject matter. Figure 3 shows a recombinant vector comprising a presently disclosed expression cassette that comprises a human β^{A-T87Q} globin gene, which is operably linked to a β -globin LCR region that comprises a 602 bp HS1 region (*e.g.*, one having the nucleotide sequence set forth in SEQ ID NO:3), a 1301bp HS3 region (*e.g.*, one having the nucleotide sequence set forth in SEQ ID NO:5), and a 446 bp HS4 region (*e.g.*, one having the nucleotide sequence set forth in SEQ ID NO:8).

[0135] Figure 4 shows one exemplary recombinant vector comprising an expression cassette in accordance with one embodiment of the presently disclosed subject matter. Figure 4 shows a recombinant vector comprising a presently disclosed expression cassette that comprises a human β^{A-T87Q} globin gene, which is operably linked to a β -globin LCR region that comprises a 1301bp HS3 region (*e.g.*, one having the nucleotide sequence set forth in SEQ ID NO:5), and a 1065 bp HS4 region (*e.g.*, one having the nucleotide sequence set forth in SEQ ID NO:6). The expression cassette shown in Figure 4 also comprises the following five erythroid-specific enhancers (shown as "EE5" in Figure 4): one erythroid-specific enhancer having the nucleotide sequence set forth in SEQ ID NO:13, one erythroid-specific enhancer having the nucleotide sequence set forth in SEQ ID NO:14, one erythroid-specific enhancer having the nucleotide sequence set forth in SEQ ID NO:15, one erythroid-specific enhancer having the nucleotide sequence set forth in SEQ ID NO:16, and one erythroid-specific enhancer having the nucleotide sequence set forth in SEQ ID NO:17.

[0136] As shown in Figures 1-4, each of the expression cassettes comprise an insulator having the nucleotide sequence set forth in SEQ ID NO:1 (*i.e.*, insulator A1). In addition, as shown in Figures 1-4, each of the expression cassettes comprise a 879 bp human β -globin 3' enhancer, which is positioned upstream of the human β -globin gene. Furthermore, as shown in Figures 1-4, each of the recombinant vectors comprise a Woodchuck hepatitis post-regulatory element (WPRE) and a bovine growth hormone polyadenylation signal in the 3' long terminal repeat (LTR) of the vector (*e.g.*, 3' to the R region in the 3' LTR).

III. Vectors, Nucleases and CRISPR-Cas Systems

[0137] The presently disclosed subject matter provides vectors and delivery systems (e.g., a non-naturally occurring or engineered nucleases or a CRISPR-Cas system) comprising the above-described expression cassettes. The vectors and delivery systems are suitable delivery vehicles for the stable introduction of globin gene (e.g., human β -globin) into the genome of a broad range of target cells to increase expression of the globin protein (human β -globin protein) in the cell.

[0138] In certain embodiments, the vector is a retroviral vector (e.g., gamma retroviral or lentiviral) that is employed for the introduction or transduction of the above-described expression cassette into the genome of a host cell (e.g., a hematopoietic stem cell, an embryonic stem cell, an induced pluripotent stem cell, or a hemogenic endothelium cell). In certain embodiments, the retroviral vector comprises an expression cassette that comprises one of the above-described insulators, e.g., insulator A1. The insulator can be positioned at the 3' or the 5' end of the expression cassette. In one embodiment, the insulator is positioned at the 3' end of the expression cassette. During reverse transcription and vector integration, the insulator positioned at the 3' end is copied into the 5' end of the expression cassette. The resulting topology places copies of the insulator between the genomic regions located at the 5' LTR and the 3' LTR of the integrated virus and enhancer activity from the 5' LTR and internal package promoter, but does not contain the enhancer in the 3' LTR. This topology can decrease genotoxicity, thereby resulting in decreased tumor formation and increased survival of the animals.

[0139] In certain embodiments, the recombinant vector further comprises a Woodchuck hepatitis post-regulatory element (WPPE) in the 3' long terminal repeat (LTR) of the vector (e.g., 3' to the R region in the 3' LTR of the vector). In certain embodiments, the recombinant vector further comprises a bovine growth hormone polyadenylation signal in addition to the WPPE in the 3' long terminal repeat (LTR) of the vector (e.g., 3' to the R region in the 3' LTR of the vector). An essential feature of therapeutic globin vectors is to achieve a high titer, sufficient for effective transduction of patient cells. By virtue of their large cargo, comprising a gene, promoter, enhancers and/or LCR elements, globin lentiviral vectors inherently have low titer, complicating their manufacture and limiting their clinical use. This problem is further compounded by the incorporation of additional genomic elements such as an insulator, which further increase the size of the vector. The WPPE can increase the titer of the recombinant vector. Addition of a bovine growth hormone polyadenylation signal to the WPPE can further increase the titer of the recombinant vector. In certain embodiments, the WPPE and the bovine growth hormone polyadenylation signal are not comprised within the expression cassette, and thus, not transferred to the cells transduced with the recombinant vector. The incorporation of these elements for enhancing the production of globin lentiviral vectors is essential to yield higher titers and hence for the clinical usefulness of the vectors described in this application.

[0140] In one non-limiting example, a presently disclosed expression cassette can be cloned into a retroviral vector and expression can be driven from its endogenous promoter, from the retroviral long terminal repeat, or from an alternative internal promoter. Combinations of retroviral vector and an appropriate packaging line are also suitable, where the capsid proteins will be functional for infecting human cells. Various amphotropic virus-producing cell lines are known, including, but not limited to, PA12 (Miller, et al. (1985) Mol. Cell. Biol. 5:431-437); PA317 (Miller, et al. (1986) Mol. Cell. Biol. 6:2895-2902); and CRIP (Danos, et al. (1988) Proc. Natl. Acad. Sci. USA 85:6460-6464). Non-amphotropic particles are suitable too, e.g., particles pseudotyped with VSVG, RD114 or GALV envelope and any other known in the art.

[0141] Suitable methods of transduction also include direct co-culture of the cells with producer cells, e.g., by the method of Bregni, et al. (1992) Blood 80:1418-1422, or culturing with viral supernatant alone or concentrated vector stocks with or without appropriate growth factors and polycations, e.g., by the method of Xu, et al. (1994) Exp. Hemat. 22:223-230; and Hughes, et al. (1992) J. Clin. Invest. 89:1817.

[0142] Transducing viral vectors can be used to express a globin gene (e.g., a human β -globin gene) in a host cell (e.g., hematopoietic stem cells, an embryonic stem cell, or an induced pluripotent stem cell). Preferably, the chosen vector exhibits high efficiency of infection and stable integration and expression (see, e.g., Cayouette et al., Human Gene Therapy (1997);8:423-430; Kido et al., Current Eye Research (1996);15:833-844; Bloomer et al., Journal of Virology (1997);71:6641-6649; Naldini et al., Science (1996);272:263-267; and Miyoshi et al., Proc. Natl. Acad. Sci. U.S.A. 94:10319, 1997). Other viral vectors that can be used include, for example, adenoviral, lentiviral, and adeno-associated viral vectors, vaccinia virus, a bovine papilloma virus, or a herpes virus, such as Epstein-Barr Virus (also see, for example, the vectors of Miller, Human Gene Therapy (1990);15-14; Friedman, Science (1989);244:1275-1281; Eglitis et al., Bio-Techniques 6:608-614, 1988; Tolstoshev et al., Current Opinion in Biotechnology (1990);1:55-61; Sharp, The Lancet (1991);337:1277-1278; Cornetta et al., Nucleic Acid Research and Molecular Biology (1987)36:311-322; Anderson, Science (1984);226:401-409; Moen, Blood Cells (1991);17:407-416; Miller et al., Biotechnology (1989);7:980-990; Le Gal La Salle et al., Science (1993);259:988-990; and Johnson, Chest (1995);107:77S-83S). Retroviral vectors are particularly well developed and have been used in clinical settings (Rosenberg et al., N. Engl. J. Med (1990);323:370; Anderson et al., U.S. Pat. No. 5,399,346).

[0143] The requirement for efficient delivery and integration make retroviral vectors suitable for transducing a presently disclosed expression cassette. Retroviral vectors can be derived from three genera of the retroviridae: the γ -retroviruses

(also known as C-type murine retroviruses or oncoretroviruses), the lentiviruses, and the spumaviruses (also known as foamy viruses). Several reviews detailing molecular approaches for the generation of replication-defective retroviral particles are available (Cornetta *et al.* (2005); Cockrell & Kafri (2007)). The vector itself, which encodes the therapeutic transgene or cDNA, retains the minimal viral sequences needed to enable packaging in viral particles in a packaging cell line, reverse transcription, and integration. The packaging cell expresses the necessary structural proteins and enzymes that are required to assemble an infectious recombinant particle that contains the vector sequence and the machinery needed for its reverse transcription and integration in the transduced cell.

[0144] While the manufacturing aspects of all retroviral vector types follow the same general principles, γ -retroviral, lentiviral and spumaviral vectors differ in some of their intrinsic biological properties. Gamma-retroviruses, including the prototypic murine leukaemia viruses (MLV), effectively infect many cell types but are unable to integrate in cells that do not proceed to S phase soon after infection. In contrast, lentiviruses and their vector derivatives can transduce nondividing cells (Follenzi & Naldini, 2002; Salmon & Trono, 2002) owing to their ability to translocate to the nucleus and integrate in the absence of cell division (Lewis & Emerman, 1994; Goff, 2001). Another fundamental attribute of lentiviral vectors is their relative genomic stability, as established for globin lentiviral vectors (May *et al.*, 2000), which contrasts with the genomic instability of MLV-based globin vectors (Leboulch *et al.*, 1994; Sadelain *et al.*, 1995). Lentiviral and foamy vectors further provide a greater packaging capacity (Kumar *et al.*, 2001; Rethwilm, 2007). All three vector types have been used successfully for the transduction of cytokine-activated HSCs (Miyoshi *et al.*, 1999; Josephson *et al.*, 2002; Leurs *et al.*, 2003).

[0145] These three vector systems differ in their integration patterns. The integration pattern of retroviruses is semi-random and biased towards genes and their vicinity in approximately two-thirds of all integration events (Schroder *et al.*, 2002; Wu *et al.*, 2003; Mitchell *et al.*, 2004; De Palma *et al.*, 2005; Trobridge *et al.*, 2006). There are however subtle and possibly significant differences in their exact distribution. Gamma-retroviruses have a propensity to integrate upstream of transcribed genes, whereas lentiviruses and lentiviral vectors target the entire transcribed gene sequence. Foamy vectors appear to be less prone to intragenic integration (Trobridge *et al.*, 2006). In one embodiment, the vector comprising the expression cassette is a lentivirus vector. The vectors can be derived from human immunodeficiency-1 (HIV-1), human immunodeficiency-2 (HIV-2), simian immunodeficiency virus (SIV), feline immunodeficiency virus (FIV), bovine immunodeficiency virus (BIV), Jembrana Disease Virus (JDV), equine infectious anemia virus (EIAV), caprine arthritis encephalitis virus (CAEV) and the like. In one non-limiting embodiment, the lentiviral vector is an HIV vector. HIV-based constructs are the most efficient at transduction of human cells.

[0146] The semi-random pattern of vector integration exposes patients to the risk of insertional oncogenesis when the vector trans-activates a neighboring oncogene. This may result in clonal expansion (Ott *et al.*, 2006; Cavazzana-Calvo *et al.*, 2010), myelodysplasia (Stein *et al.*, 2010) or leukaemia (Hacein-Bey-Abina *et al.*, 2003, 2008; Howe *et al.*, 2008). Targeted gene delivery strategies, utilizing a non-naturally occurring or engineered nuclease (including, but not limited to, Zinc-finger nuclease (ZNFs), meganuclease, transcription activator-like effector nuclease (TALEN)), or a CRISPR-Cas system, can reduce or even eliminate the concern of insertional oncogenesis that is inherent to the use of retroviral vectors.

[0147] Eukaryotic cells utilize two distinct DNA repair mechanisms in response to DNA double strand breaks (DSBs): Homologous recombination (HR) and non-homologous end-joining (NHEJ). The activation of the HR repair machinery depends on the cell cycle status, and it is restricted to the S and G2 phases; in contrast, the NHEJ pathway is active throughout the cell cycle. Mechanistically, HR is an error-free DNA repair mechanism, because it requires a homologous template to repair the damaged DNA strand. On the other hand, NHEJ is a template-independent repair mechanism that is imprecise, due to DNA end processing during repair that leads to insertions or deletions at the DNA break site (Moynahan & Jasin, 2010). Because of its homology-based mechanism, HR has been used as a tool to site-specifically engineer the genome of different species. From a therapeutic perspective, HR has been successfully used to repair mutated genes, thus offering a promising approach to cell-mediated treatment of monogenic diseases (Porteus *et al.*, 2006).

[0148] Gene targeting by HR requires the use of two homology arms that flank the transgene/target site of interest. Generally, standard plasmid DNAs have been used to deliver 5-10 kb homology arms along with transgenes for positive and negative selection. This method is commonly used to knock-out/knockin genes in mouse embryonic stem (mES) cells (Capecchi, 2005; Fig 2B). In human cells, the use of this approach has allowed gene targeting with efficiencies in the order of 10^{-6} , which are lower than in mES cells and are not therapeutically practical. HR efficiency can be increased by the introduction of DNA-doubled stranded breaks (DSBs) at the target site using specific rare-cutting endonucleases, resulting in over 1,000-fold increase in correct gene targeting (Jasin, 1996). The discovery of this phenomenon prompted the development of methods to create site-specific DSBs in the genome of different species. Various chimeric enzymes have been designed for this purpose over the last decade, namely zinc-finger nucleases (ZFNs), meganucleases, and transcription activator-like effector nucleases (TALENs).

[0149] ZFNs are modular chimeric proteins that contain a ZF-based DNA binding domain (DBD) and a FokI nuclease domain (Porteus & Carroll, 2005). DBD is usually composed of three ZF domains, each with 3-base pair specificity; the

FokI nuclease domain provides a DNA nicking activity, which is targeted by two flanking ZFNs. Owing to the modular nature of the DBD, any site in a genome could be targeted in principle. However, as a single ZFN can bind and nick DNA, there is potential for a high number of off-target effects, resulting in the activation of the NHEJ pathway that may either introduce insertions/deletions or integrate the targeting vector in a non-specific manner. Obligate FokI domains that can nick their respective DNA strand only when they form a heterodimer were recently reported (Doyon et al, 2011). The use of such obligate ZFNs can reduce the genotoxic effects of this approach.

[0150] Meganucleases (MNs)/homing endonucleases (HEs) are dsDNA nucleases that recognize and cleave large DNA sites (14-40 bp) with low cleavage frequencies in eukaryotic genomes (Paques & Duchateau, 2007). Although this limits the potential target sites, MN-DNA structures have been used as a guide to specifically modify DNA-interacting residues in order to change the MN specificity (Marcaida et al, 2010). I-CreI has been successfully engineered to generate chimeric meganucleases that target the human XPC and RAG1 genes, and they have been shown to stimulate HR activity in mammalian cells with no evident genotoxicity (Redondo et al, 2008; Grizot et al, 2009). The genotoxicity of this approach will need to be compared to that of ZFNs and TALE nucleases.

[0151] TALENs are similar ZFN except that the DBD is derived from transcription activator-like effectors (TALEs), which are virulent factors used by phytopathogenic bacteria (Herbers, 1992). The TALE DBD is modular, and it is composed of 34-residue repeats, and its DNA specificity is determined by the number and order of repeats (Herbers, 1992). Each repeat binds a single nucleotide in the target sequence through only two residues (Boch, 2011). The advantage over ZFN technology is the rapid construction of DBDs.

[0152] A number of studies have used these chimeric enzymes to stimulate HR for either gene addition or gene repair at their target site (Paques & Duchateau, 2007; Urnov et al, 2010). Porteus designed a ZFN to a half site sequence from the human HBB that surrounds the sickle cell mutation nucleotide (Porteus, 2006). This ZFN targets the sequence and stimulates HR at a chimeric DNA target when combined with a ZFN targeting the Zif268 binding site. There have been recent advances in targeting genes in cord blood CD34⁺ cells. Use of non-integrating lentiviruses to deliver ZFNs and the donor DNA in these cells to target the CCR5 gene was reported in Lombardo et al, 2007. Lombardo et al, 2007 showed gene addition at this locus with correct targeting in 80% of the positively selected cells.

[0153] The presently disclosed subject matter provides a non-naturally occurring or engineered nuclease comprising a presently disclosed expression cassette, as described above. Suitable nucleases include, but are not limited to, ZFNs, meganucleases, and TALENs. A presently disclosed nuclease comprises a DNA binding domain and a nuclease cleavage domain. The DNA binding domain of the nuclease can be engineered to bind to a sequence of choice, *e.g.*, a predetermined site. An engineered DNA binding domain can have a distinct binding specificity, compared to a naturally occurring nuclease. Engineering methods include, but are not limited to, rational design and various types of selection. Any suitable cleavage domain can be operatively linked to a DNA-binding domain to form a nuclease. For example, Zinc-finger protein (ZFP) DNA-binding domains can be fused to nuclease cleavage domains to create ZFNs—a functional entity that is able to recognize its intended nucleic acid target through its engineered ZFP DNA binding domain and cause the DNA to be cut near the ZFP binding site via the nuclease activity. *See, e.g.*, Kim et al. Proc Nat'l Acad Sci USA (1996);93(3):1156-1160. Likewise, TALE DNA-binding domains can be fused to nuclease cleavage domains to create TALENs. *See, e.g.*, U.S. Publication No. 20110301073.

[0154] The cleavage domain can be heterologous to the DNA-binding domain, *e.g.*, a meganuclease DNA-binding domain and cleavage domain from a different nuclease. Heterologous cleavage domains can be obtained from any endonuclease or exonuclease. Exemplary endonucleases from which a cleavage domain can be derived include, but are not limited to, restriction endonucleases and homing endonucleases. *See, for example*, 2002-2003 Catalog, New England Biolabs, Beverly, Mass.; and Belfort et al. (1997) Nucleic Acids Res.25:3379-3388. Additional enzymes which cleave DNA are known (*e.g.*, S1 Nuclease; mung bean nuclease; pancreatic DNase I; micrococcal nuclease; yeast HO endonuclease; *see also* Linn et al. (eds.) Nucleases, Cold Spring Harbor Laboratory Press, 1993). One or more of these enzymes (or functional regions thereof) can be used as a source of cleavage domains and cleavage half-domains.

[0155] Similarly, a cleavage half-domain can be derived from the above-described nuclease that requires dimerization for cleavage activity. In general, two fusion proteins are required for cleavage if the fusion proteins comprise cleavage half-domains. Alternatively, a single protein comprising two cleavage half-domains can be used. The two cleavage half-domains can be derived from the same endonuclease (or functional portions thereof), or each cleavage half-domain can be derived from a different endonuclease (or functional portions thereof).

[0156] In certain embodiments, the nuclease comprises an expression cassette that comprises two of the above-described insulators, *e.g.*, two of the insulator having the nucleotide sequence set forth in SEQ ID NO:1. One of the two insulators is positioned at the 3' end of the expression cassette, and the other insulator is positioned at the 5' end of the expression cassette.

[0157] The presently disclosed subject matter also provides a non-naturally occurring or engineered CRISPR-Cas system comprising the above-described expression cassette. The CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas (CRISPR Associated) system is an engineered nuclease system based on a bacterial system that can be used for genome engineering. It is based on part of the adaptive immune response of many bacteria and archaea.

When a virus or plasmid invades a bacterium, segments of the invader's DNA are converted into CRISPR RNAs (crRNA) by the "immune" response. The crRNA then associates, through a region of partial complementarity, with another type of RNA called tracrRNA to guide a CRISPR-Cas nuclease to a region homologous to the crRNA in the target DNA called a "proto spacer". The CRISPR-Cas nuclease cleaves the DNA to generate blunt ends at the DSB at sites specified by a 20-nucleotide guide sequence contained within the crRNA transcript. The CRISPR-Cas nuclease requires both the crRNA and the tracrRNA for site specific DNA recognition and cleavage. This system has been engineered such that the crRNA and tracrRNA can be combined into one molecule (the "single guide RNA"); and the crRNA equivalent portion of the single guide RNA can be engineered to guide the CRISPR-Cas nuclease to target any desired sequence (see Jinek et al., *Science* (2012);337:816-821). Thus, the CRISPR-Cas system can be engineered to create a DSB at a desired target in a genome. In certain embodiments, the CRISPR-Cas system comprises a CRISPR-Cas nuclease and a single-guide RNA. Suitable examples of CRISPR-Cas nucleases include, but are not limited to, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homologs thereof, or modified versions thereof. These CRISPR-Cas nucleases are known; for example, the amino acid sequence of *S. pyogenes* Cas9 protein may be found in the SwissProt database under accession number Q99ZW2. In some embodiments, the CRISPR-Cas nuclease has DNA cleavage activity, e.g., Cas9. In certain embodiments, the CRISPR-Cas nuclease is Cas9. The CRISPR-Cas nuclease can direct cleavage of one or both strands at the location of a target sequence (e.g., a genomic safe harbor site). Additionally, the CRISPR-Cas nuclease can direct cleavage of one or both strands within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, or more base pairs from the first or last nucleotide of a target sequence.

[0158] The presently disclosed nucleases and CRISPR-Cas system allow for targeted delivery of the expression cassette. In certain embodiments, a presently disclosed CRISPR-Cas system or the DNA binding domain of a presently disclosed nuclease binds to a genomic safe harbor site. A nuclease or the CRISPR-Cas system generates a double strand break at the genomic safe harbor site. Genomic safe harbor sites are intragenic or extragenic regions of the human genome that are able to accommodate the predictable expression of newly integrated DNA without adverse effects on the host cell or organism. A useful safe harbor must permit sufficient transgene expression to yield desired levels of the vector-encoded protein or non-coding RNA. A genomic safe harbor site also must not predispose cells to malignant transformation nor alter cellular functions. Methods for identifying genomic safe harbor sites are described in Sadelain et al., "Safe Harbours for the integration of new DNA in the human genome," *Nature Reviews* (2012);12:51-58; Papapetrou et al., "Genomic safe harbors permit high β -globin transgene expression in thalassemia induced pluripotent stem cells" *Nat Biotechnol.* (2011) Jan;29(1):73-8, which are incorporated by reference in their entireties. A presently disclosed genomic safe harbor site meets one or more (one, two, three, four, or five) of the following five criteria: (i) distance of at least 50 kb from the 5' end of any gene (e.g., from the 5' end of the gene), (ii) distance of at least 300 kb from any cancer-related gene, (iii) within an open/accessible chromatin structure (measured by DNA cleavage with natural or engineered nucleases), (iv) location outside a gene transcription unit and (v) location outside ultraconserved regions (UCRs), microRNA or long non-coding RNA of the human genome. As the most common insertional oncogenesis event is transactivation of neighboring tumor-promoting genes, the first two criteria exclude the portion of the human genome located near promoters of genes, in particular, cancer-related genes, which are genes functionally implicated in human cancers or the human homologs of genes implicated in cancer in model organisms. Proximity to miRNA genes is one exclusion criterion because miRNAs are implicated in the regulation of many cellular processes, including cell proliferation and differentiation. As vector integration within a transcription unit can disrupt gene function through the loss of function of a tumor suppressor gene or the generation of an aberrantly spliced gene product, the fourth (iv) criterion excludes all sites located inside transcribed genes. UCRs, which are regions that are highly conserved over multiple vertebrates and known to be enriched for enhancers and exons, and long non-coding RNAs, are also excluded. In certain embodiments, the genomic safe harbor site is an extragenic genomic safe harbor site. In certain embodiments, the genomic safe harbor site is located on chromosome 1.

[0159] The presently disclosed subject matter also provides polynucleotides encoding the above-described nucleases, vectors comprising the polynucleotides encoding the above-described nucleases, polynucleotides encoding the above-described CRISPR-Cas system, and vectors comprising the polynucleotides encoding the above-described CRISPR-Cas system.

[0160] The nucleases and polynucleotides encoding these nucleases, and the CRISPR-Cas system and polynucleotides encoding the CRISPR-Cas system can be delivered *in vivo* or *ex vivo* by any suitable means. For example, nucleases and CRISPR-Cas system as described herein can be delivered to a cell (e.g., a hematopoietic stem cell, an embryonic stem cell, an induced pluripotent stem cell, or an hemogenic endothelium cell) by a vector comprising polynucleotides encoding the nuclease or the CRISPR-Cas system. Any vectors can be used including, but not limited to, plasmid vectors, retroviral vectors (e.g., γ -retroviral vectors, lentiviral vectors and foamy viral vectors), adenovirus vectors, poxvirus vectors; herpes virus vectors and adena-associated virus vectors, etc. In one embodiment, the vector comprising a polynucleotide encoding an above-described nuclease or an above-described CRISPR-Cas system is a lentiviral

vector. In one particular embodiment, the lentiviral vector is a non-integrating lentiviral vector. Examples of non-integrating lentiviral vector are described in Ory et al. (1996) Proc. Natl. Acad. Sci. USA 93:11382-11388; Dull et al., (1998) J. Viral. 72:8463-8471; Zuffery et al. (1998) J. Viral. 72:9873-9880; Follenzi et al., (2000) Nature Genetics 25:217-222; U.S. Patent Publication No 2009/054985.

[0161] Additionally, non-viral approaches can also be employed for the expression of a globin gene in cells. For example, a nucleic acid molecule can be introduced into a cell by administering the nucleic acid in the presence of lipofection (Feigner et al., Proc. Natl. Acad. Sci. U.S.A. 84:7413, 1987; Ono et al., Neuroscience Letters 17:259, 1990; Brigham et al., Am. J. Med. Sci. 298:278, 1989; Staubinger et al., Methods in Enzymology 101:512, 1983), asialoorosomucoid-polylysine conjugation (Wu et al., Journal of Biological Chemistry 263:14621, 1988; Wu et al., Journal of Biological Chemistry 264:16985, 1989), or by micro-injection under surgical conditions (Wolff et al., Science 247:1465, 1990). Other non-viral means for gene transfer include transfection *in vitro* using calcium phosphate, DEAE dextran, electroporation, and protoplast fusion. Liposomes can also be potentially beneficial for delivery of DNA into a cell. Transplantation of normal genes into the affected tissues of a subject can also be accomplished by transferring a normal nucleic acid into a cultivatable cell type *ex vivo* (e.g., an autologous or heterologous primary cell or progeny thereof), after which the cell (or its descendants) are injected into a targeted tissue or are injected systemically. Recombinant receptors can also be derived or obtained using transposases. Transient expression may be obtained by RNA electroporation.

IV. Cells

[0162] Genetic modification of cells (e.g., hematopoietic stem cells, embryonic stem cells, induced pluripotent stem cells, and hemogenic endothelium cells) can be accomplished by transducing a substantially homogeneous cell composition with a recombinant DNA or RNA construct (e.g., a vector or a delivery system comprising the above-described expression cassette). The presently disclosed subject matter provides cells transduced with the above-described expression cassettes, cells transduced with the above-described vectors, and cells transduced with the above-described nucleases or with vectors comprising polynucleotides encoding the nucleases, and cell transduced with the above-described CARISPR-Cas system or with vectors comprising polynucleotides encoding the CARISPR-Cas system, which are collectively referred to as "transduced cells". As described above, the vectors, nucleases and CRISPR-Cas system are employed for transduction of the expression cassette to the cells to express a globin gene (e.g., a human β -globin gene). In certain embodiments, the transduced cells are administered to a subject to treat and/or prevent a hematopoietic disease, disorder, or condition. The presently disclosed insulators can enhance the efficiency of the transduction of the the expression cassette to cells.

[0163] Suitable transduced cells include, but are not limited to, stem cells, progenitor cells, and differentiated cells. As used herein, the term "progenitor" or "progenitor cells" refers to cells that have the capacity to self-renew and to differentiate into more mature cells. Progenitor cells have a reduced potency compared to pluripotent and multipotent stem cells. Many progenitor cells differentiate along a single lineage, but may also have quite extensive proliferative capacity.

[0164] In certain embodiments, the transduced cells are stem cells. Stem cells have the ability to differentiate into the appropriate cell types when administered to a particular biological niche, *in vivo*. A stem cell is an undifferentiated cell capable of (1) long term self-renewal, or the ability to generate at least one identical copy of the original cell, (2) differentiation at the single cell level into multiple, and in some instance only one, specialized cell type and (3) of *in vivo* functional regeneration of tissues. Stem cells are sub-classified according to their developmental potential as totipotent, pluripotent, multipotent and oligo/unipotent. As used herein, the term "pluripotent" means the ability of a cell to form all lineages of the body or soma (i.e., the embryo proper). For example, embryonic stem cells are a type of pluripotent stem cells that are able to form cells from each of the three germ layers, the ectoderm, the mesoderm, and the endoderm. As used herein, the term "multipotent" refers to the ability of an adult stem cell to form multiple cell types of one lineage. For example, hematopoietic stem cells are capable of forming all cells of the blood cell lineage, e.g., lymphoid and myeloid cells.

[0165] In certain embodiments, the transduced cells are embryonic stem cells, bone marrow stem cells, umbilical cord stem cells, placental stem cells, mesenchymal stem cells, neural stem cells, liver stem cells, pancreatic stem cells, cardiac stem cells, kidney stem cells, and/or hematopoietic stem cells. In one embodiment, the transduced cells are hematopoietic stem cells (HSCs). HSCs give rise to committed hematopoietic progenitor cells (HPCs) that are capable of generating the entire repertoire of mature blood cells over the lifetime of an organism. The term "hematopoietic stem cell" or "HSC" refers to multipotent stem cells that give rise to all blood cell types of an organism, including myeloid (e.g., monocytes and macrophages, neutrophils, basophils, eosinophils, erythrocytes, megakaryocytes/platelets, dendritic cells), and lymphoid lineages (e.g., T-cells, B-cells, NK-cells). When transplanted into lethally irradiated animals or humans, hematopoietic stem and progenitor cells can repopulate the erythroid, neutrophil-macrophage, megakaryocyte and lymphoid hematopoietic cell pool.

[0166] HSCs can be isolated or collected from bone marrow, umbilical cord blood, or peripheral blood. HSCs can be identified according to certain phenotypic or genotypic markers. For example, HSCs can be identified by their small size, lack of lineage (lin) markers, low staining (side population) with vital dyes such as rhodamine 123 (rhodamineDULL, also called rholo) or Hoechst 33342, and presence of various antigenic markers on their surface, many of which belong to the cluster of differentiation series (e.g., CD34, CD38, CD90, CD133, CD105, CD45, Ter119, and c-kit, the receptor for stem cell factor). In one embodiment, the transduced cell is a CD34⁺ HSC.

[0167] In one embodiment, the transduced cell is an embryonic stem cell. In another embodiment, the transduced cell is an induced pluripotent stem cell. In yet another embodiment, the transduced cell is a hemogenic endothelium cell.

[0168] While HSCs are the natural vehicle for restoring long-term hematopoiesis, their use has some important limitations. The first is their relative scarcity, which can eventually preclude autologous HSC therapy when the harvested cellular product is too small. The second is the difficulty to perform biosafety testing such as integration site analysis and consequently to select cells with chosen integration sites, because adult HSCs cannot be replicated *in vitro*. The third limitation is that homologous recombination using current technologies is practically impossible thus compromising the advent of gene correction. All of these limitations are ultimately due to the fact that adult HSCs cannot be expanded *in vitro* without losing their stem cell potency. These limitations explain the critical importance of viral vectors such as gamma-retroviral and lentiviral vectors, which are remarkably quick and efficient in achieving stable gene transfer. This is essential when dealing with HSCs that are only available in limited quantities.

[0169] Use of ESs and induced pluripotent stem (iPS) cells for globin gene therapy is disclosed in Moi et al., Haematol March 1, 2008;93(3):325-330. Embryonic stem (ES) cells are amenable to gene targeting and correction, which requires unlimited *in vitro* cell division without losing multipotency. Chang et al., Proc Natl Acad Sci USA 2006;103:1036-40 provided proof of principle of the feasibility of the homologous recombination approach in mice with sickle cell anemia. Takahashi et al. Cell 2006;126:663-76 reported the successful reprogramming of fibroblasts to an embryonic stem-like state. Cells obtained by this reverse-differentiation process, called induced pluripotent stem (iPS) cells, were produced by exposing embryonic or young adult bulk fibroblast cultures to gamma-retroviral vectors encoding 4 transcription factors, which are physiologically active in the embryonic stem cells, but generally turned off when differentiation progresses. The cultured cells formed colonies similar to ES cell colonies. These findings have been confirmed and extended by others to both mouse and human fibroblasts (Meissner et al., Nat Biotechnol 2007;25:1177-81; Nakagawa et al., Nat Biotechnol 2007;26:101-6; Okita et al., Nature 2007; 448:313-7; Park et al., Nature 2007;451:141-6; Takahashi et al., Nat Protoc 2007;2:3081-9; Takahashi K et al., Cell 2007;131:861-72; Wernig et al., Nature 2007;448:318-24; Yu J et al., Science 2007;318: 1917-20). Rudolf Jaenisch and co-workers achieved a successful gene therapy in a mouse model of sickle cell disease, using homologous recombination in ES-like iPS cells (Hanna et al., Science 2007;318:1920-3). The process has so far been mostly applied to fibroblast harvested from a skin biopsy, which are then induced to become iPS by transduction with retroviral vectors that encode four stem cell transcription factors. iPS are amenable to the correction of the SC mutation by standard homologous recombination techniques and can then be differentiated *in vitro* into unlimited amounts of hematopoietic stem cells. The whole process ends with the autologous transplantation of the corrected HSC into the original mouse donor, which will now be cured of its SC disease. This technique is not only useful for homologous recombination, but can also enhance lentiviral-mediated globin gene transfer for the treatment of β -thalassemia by providing a means to perform detailed integration site analysis and adequate *in vitro* cell expansion before infusing cells into the recipient.

[0170] The cell of the presently disclosed subject matter can be autologous ("self") or non-autologous ("non-self," e.g., allogeneic, syngeneic or xenogeneic). As used herein, "autologous" refers to cells from the same subject. As used herein, "allogeneic" refers to cells of the same species that differ genetically to the cell in comparison. As used herein, "syngeneic" refers to cells of a different subject that are genetically identical to the cell in comparison. As used herein, "xenogeneic" refers to cells of a different species to the cell in comparison. In certain embodiments, the cell is autologous, e.g., a cell transduced with the presently disclosed expression cassette is administered to a subject from whom the cell is collected, e.g., the cell is collected from bone marrow, umbilical cord blood, peripheral blood, and/or adipose tissue of the subject. In certain embodiments, the cell is obtained or collected from bone marrow of a subject.

[0171] In certain embodiments, prior to transduction with the expression cassette, the cell is pre-stimulated, e.g., in the presence of one or more cytokines (e.g., IL-3, IL-1 α , IL-6, Kit ligand (also known as "Stem Cell Factor (SCF)"), and Flt-3 ligand), and/or one or more glycoproteins (e.g., thrombopoietin and fibronectin). In one non-limiting example, the cell is pre-stimulated in the presence of Flt-3 ligand, SCF, thrombopoietin, interleukin-3, and fibronectin. The cell can be pre-stimulated for about 24 hours or longer, e.g., about 48 hours, or about 36 hours. Subsequently, the cell is transduced with a presently disclosed expression cassette, or a vector or another delivery system comprising such expression cassette. Transduction can be performed on a fresh cell, or on a frozen cell. Genomic DNA of the cell is isolated to determine the vector copy number and analyze the integration site or integrated vector structure, e.g., by South blot analysis and/or by Quantitative PCR. For quantification of globin mRNA (e.g., human β -globin transgene analysis), total RNA is extracted from the cell. Quantitative primer extension assay can be used for quantification of globin mRNA.

V. Compositions and Formulations

[0172] The presently disclosed subject matter provides pharmaceutical compositions comprising a presently disclosed transduced cell as described above and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" includes any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like that are physiologically compatible, including pharmaceutically acceptable cell culture media. The pharmaceutically acceptable carrier can be suitable for parenteral (e.g., intravenous, intramuscular, subcutaneous, or intraperitoneal), spinal or epidermal administration (e.g., by injection, infusion or implantation). Depending on the route of administration, the active compound, e.g., the transduced cell, may be coated in a material to protect the compound from the action of acids and other natural conditions that may inactivate the compound.

[0173] Pharmaceutically acceptable carriers include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the transduced cells, use thereof in the pharmaceutical compositions of the invention is contemplated.

[0174] The pharmaceutical compositions of the presently disclosed subject matter can further comprise one or more polypeptides, polynucleotides, vectors comprising the same, transduced cells, etc., as described herein, formulated in pharmaceutically-acceptable or physiologically- acceptable solutions for administration to a cell or an animal, either alone, or in combination with one or more other modalities of therapy. If desired, the pharmaceutical compositions of the presently disclosed subject matter can be administered in combination with other agents, including, but not limited to, cytokines, growth factors, hormones, small molecules or various pharmaceutically-active agents. Any additional agents that do not adversely affect the ability of the composition to deliver the intended gene therapy can be included in the compositions.

[0175] In the pharmaceutical compositions of the presently disclosed subject matter, formulation of pharmaceutically-acceptable excipients and carrier solutions is well known to those of ordinary skill in the art, as is the development of suitable dosing and treatment regimens for using the particular compositions described herein in a variety of treatment regimens, including, e.g., oral, parenteral, intravenous, intranasal, and intramuscular administration and formulation.

[0176] The pharmaceutical compositions of the presently disclosed subject matter can be delivered parenterally (e.g., intravenously, intramuscularly, or intraperitoneally) as described, for example, in U.S. Pat. No. 5,543,158; U.S. Pat. No. 5,641,515 and U.S. Pat. No. 5,399,363. Solutions of the active compounds as free base or pharmacologically acceptable salts may be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose. Dispersions may also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations contain a preservative to prevent the growth of microorganisms.

[0177] Pharmaceutically acceptable carriers include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. The use of such media and agents for pharmaceutically active substances is known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the pharmaceutical compositions of the invention is contemplated. Supplementary active compounds can also be incorporated into the compositions.

[0178] Therapeutic compositions typically must be sterile and stable under the conditions of manufacture and storage. The composition can be formulated as a solution, microemulsion, liposome, or other ordered structure suitable to high drug concentration. The pharmaceutically acceptable carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, or sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent that delays absorption, for example, monostearate salts and gelatin.

[0179] The pharmaceutical compositions of the presently disclosed subject matter can be conveniently provided as sterile liquid preparations, e.g., isotonic aqueous solutions, suspensions, emulsions, dispersions, or viscous compositions, which can be buffered to a selected pH. Liquid preparations are normally easier to prepare than gels, other viscous compositions, and solid compositions. Additionally, liquid compositions are somewhat more convenient to administer, especially by injection. Viscous compositions, on the other hand, can be formulated within the appropriate viscosity range to provide longer contact periods with specific tissues. Liquid or viscous compositions can comprise carriers, which can be a solvent or dispersing medium containing, for example, water, saline, phosphate buffered saline, polyol (for example, glycerol, propylene glycol, liquid polyethylene glycol, and the like) and suitable mixtures thereof.

[0180] Sterile injectable solutions can be prepared by incorporating the compositions of the presently disclosed subject matter in the required amount of the appropriate solvent with various amounts of the other ingredients, as desired. Such compositions may be in admixture with a suitable carrier, diluent, or excipient such as sterile water, physiological saline, glucose, dextrose, or the like. The compositions can also be lyophilized. The compositions can contain auxiliary sub-

stances such as wetting, dispersing, or emulsifying agents (e.g., methylcellulose), pH buffering agents, gelling or viscosity enhancing additives, preservatives, flavoring agents, colors, and the like, depending upon the route of administration and the preparation desired. Standard texts, such as "REMINGTON'S PHARMACEUTICAL SCIENCE", 17th edition, 1985, incorporated herein by reference, may be consulted to prepare suitable preparations, without undue experimentation.

[0181] Various additives which enhance the stability and sterility of the compositions, including antimicrobial preservatives, antioxidants, chelating agents, and buffers, can be added. Prevention of the action of microorganisms can be ensured by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, and the like. Prolonged absorption of the injectable pharmaceutical form can be brought about by the use of agents delaying absorption, for example, aluminum monostearate and gelatin.

[0182] The compositions can be isotonic, i.e., they can have the same osmotic pressure as blood and lacrimal fluid. The desired isotonicity of the compositions of the presently disclosed subject matter can be accomplished using sodium chloride, or other pharmaceutically acceptable agents such as dextrose, boric acid, sodium tartrate, propylene glycol or other inorganic or organic solutes. Sodium chloride is preferred particularly for buffers containing sodium ions.

[0183] For parenteral administration in an aqueous solution, for example, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by sterilization micro filtration. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying (lyophilization) that yield a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[0184] In certain embodiments, the compositions can be delivered by intranasal sprays, inhalation, and/or other aerosol delivery vehicles. Methods for delivering genes, polynucleotides, and peptide compositions directly to the lungs via nasal aerosol sprays are described, e.g., in U.S. Pat. No. 5,756,353 and U.S. Pat. No. 5,804,212. Methods of delivering drugs using lysophosphatidyl-glycerol compounds are described, e.g., in U.S. Pat. No. 5,725,871. Transmucosal drug delivery in the form of a polytetrafluoroethylene support matrix is described, e.g., in U.S. Pat. No. 5,780,045. The compositions of the presently disclosed subject matter can be formulated for delivery either encapsulated in a lipid particle, a liposome, a vesicle, a nanosphere, a nanoparticle or the like. The formulation and use of such delivery vehicles can be carried out using known and conventional techniques. The formulations and compositions of the presently disclosed subject matter can comprise one or more repressors and/or activators comprising a combination of any number of polypeptides, polynucleotides, and small molecules, as described herein, formulated in pharmaceutically-acceptable or physiologically-acceptable solutions (e.g., culture medium) for administration to a cell or an animal, either alone, or in combination with one or more other modalities of therapy.

[0185] In certain aspects, the presently disclosed subject matter provides formulations or compositions suitable for the delivery of viral vector systems (i.e., viral-mediated transduction) including, but not limited to, retroviral (e.g., lentiviral) vectors. Exemplary formulations for *ex vivo* delivery can also include the use of various transfection agents known in the art, such as calcium phosphate, electroporation, heat shock and various liposome formulations (i.e., lipid-mediated transfection). Liposomes are lipid bilayers entrapping a fraction of aqueous fluid. DNA spontaneously associates to the external surface of cationic liposomes (by virtue of its charge) and these liposomes will interact with the cell membrane.

[0186] The skilled artisan can readily determine the amount of cells and optional additives, vehicles, and/or carrier in compositions and to be administered in methods of the presently disclosed subject matter. Typically, any additives (in addition to the transduced cell(s) and/or agent(s)) are present in an amount of from about 0.001% to about 50% by weight) solution in phosphate buffered saline, and the active ingredient is present in the order of micrograms to milligrams, such as from about 0.0001 wt% to about 5 wt %, from about 0.0001 wt% to about 1 wt %, from about 0.0001 wt% to about 0.05 wt%, from about 0.001 wt% to about 20 wt %, from about 0.01 wt% to about 10 wt %, or from about 0.05 wt% to about 5 wt %. For any composition to be administered to an animal or human, and for any particular method of administration, toxicity should be determined, such as by determining the lethal dose (LD) and LD50 in a suitable animal model e.g., rodent such as mouse; and, the dosage of the composition(s), concentration of components therein and timing of administering the composition(s), which elicit a suitable response. Such determinations do not require undue experimentation from the knowledge of the skilled artisan, this disclosure and the documents cited herein. And, the time for sequential administrations can be ascertained without undue experimentation.

VI. Uses and Methods

[0187] Vectors and other delivery systems (nucleases and CRISPR-Cas systems) comprising the presently disclosed expression cassette provide improved methods of gene therapy. As used herein, the term "gene therapy" refers to the introduction of a polynucleotide into a cell's genome that restores, corrects, or modifies the gene and/or expression of

the gene. In various non-limiting embodiments, a presently disclosed vector or other delivery system (e.g., a nuclease or a CRISPR-Cas system) comprises an expression cassette comprising a globin gene or a functional portion thereof that encodes a globin protein (e.g., human β globin protein), which provides curative, preventative, or ameliorative benefits to a subject diagnosed with or that is suspected of having a disease, disorder, or condition of the hematopoietic system. The vector or other delivery systems (e.g., a nuclease and the CRISPR-Cas system) can infect and transduce the cell *in vivo*, *ex vivo*, or *in vitro*. In *ex vivo* and *in vitro* embodiments, the transduced cells can then be administered to a subject in need of therapy. The presently disclosed subject matter contemplates that the vectors and other delivery systems (e.g., nucleases or CRISPR-Cas systems), viral particles, and transduced cells of the presently disclosed subject matter are be used to treat, prevent, and/or ameliorate a disease, disorder, or condition of the hematopoietic system in a subject, e.g., a hemoglobinopathy.

[0188] As used herein, the term "hemoglobinopathy" or "hemoglobinopathic condition" includes any disorder involving the presence of an abnormal hemoglobin molecule in the blood. Examples of hemoglobinopathies included, but are not limited to, hemoglobin C disease, hemoglobin sickle cell disease (SCD), sickle cell anemia, and thalassemias. Also included are hemoglobinopathies in which a combination of abnormal hemoglobins are present in the blood (e.g., sickle cell/Hb-C disease).

[0189] As used herein, "thalassemia" refers to a hereditary disorder characterized by defective production of hemoglobin. Examples of thalassemias include α - and β -thalassemia. β -thalassemias are caused by a mutation in the beta globin chain, and can occur in a major or minor form. In the major form of β -thalassemia, children are normal at birth, but develop anemia during the first year of life. The mild form of β -thalassemia produces small red blood cells and the thalassemias are caused by deletion of a gene or genes from the globin chain. α -thalassemia typically results from deletions involving the HBA1 and HBA2 genes. Both of these genes encode α -globin, which is a component (subunit) of hemoglobin. There are two copies of the HBA1 gene and two copies of the HBA2 gene in each cellular genome. As a result, there are four alleles that produce α -globin. The different types of a thalassemia result from the loss of some or all of these alleles. Hb Bart syndrome, the most severe form of a thalassemia, results from the loss of all four α -globin alleles. HbH disease is caused by a loss of three of the four [alpha]-globin alleles. In these two conditions, a shortage of [alpha]-globin prevents cells from making normal hemoglobin. Instead, cells produce abnormal forms of hemoglobin called hemoglobin Bart (Hb Bart) or hemoglobin H (HbH). These abnormal hemoglobin molecules cannot effectively carry oxygen to the body's tissues. The substitution of Hb Bart or HbH for normal hemoglobin causes anemia and the other serious health problems associated with a thalassemia.

[0190] As used herein, the term "sickle cell disease" refers to a group of autosomal recessive genetic blood disorders, which results from mutations in a globin gene and which is characterized by red blood cells that assume an abnormal, rigid, sickle shape. They are defined by the presence of β^S -gene coding for a β -globin chain variant in which glutamic acid is substituted by valine at amino acid position 6 of the peptide, and second β -gene that has a mutation that allows for the crystallization of HbS leading to a clinical phenotype. As used herein, the term "sickle cell anemia" refers to a specific form of sickle cell disease in patients who are homozygous for the mutation that causes HbS. Other common forms of sickle cell disease include HbS/ β -thalassemia, HbS/HbC and HbS/HbD.

[0191] In certain embodiments, gene therapy methods of the presently disclosed subject matter are used to treat, prevent, or ameliorate a hemoglobinopathy that is selected from the group consisting of: hemoglobin C disease, hemoglobin sickle cell disease (SCD), sickle cell anemia, hereditary anemia, thalassemia, β -thalassemia, thalassemia major, thalassemia intermedia, α -thalassemia, and hemoglobin H disease. In one non-limiting embodiment, the hemoglobinopathy is β -thalassemia. In another non-limiting embodiment, the hemoglobinopathy is sickle cell anemia.

[0192] In various non-limiting embodiments, vectors or other delivery systems (e.g., nucleases or CRISPR-Cas systems) comprising a presently disclosed expression cassette are administered by direct injection to a cell, tissue, or organ of a subject in need of gene therapy, *in vivo*. In various other embodiments, cells are transduced *in vitro* or *ex vivo* with vectors or other delivery systems (e.g., nucleases or CRISPR-Cas systems) of the presently disclosed subject matter, and optionally expanded *ex vivo*. The transduced cells are then administered to a subject in need of gene therapy, e.g., within a pharmaceutical formulation disclosed herein.

[0193] The presently disclosed subject matter provides a method of providing a transduced cell to a subject. In various non-limiting embodiments, the method comprises administering (e.g., parenterally) one or more cells (a population of cells) transduced with a presently disclosed expression cassette or a vector or another delivery system (e.g., a nuclease or CRISPR-Cas system) comprising such expression cassette to the subject.

[0194] The presently disclosed subject matter provides a method of treating a hemoglobinopathy in a subject. In various non-limiting embodiments, the method comprises administering an effective amount of a presently disclosed transduced cell or a population of the presently disclosed transduced cells (e.g., HSCs, embryonic stem cells, or iPSCs) to the subject.

[0195] For treatment, the amount administered is an amount effective in producing the desired effect. An effective amount can be provided in one or a series of administrations. An effective amount can be provided in a bolus or by continuous perfusion. An "effective amount" (or "therapeutically effective amount") is an amount sufficient to affect a

beneficial or desired clinical result upon treatment. An effective amount can be administered to a subject in one or more doses. In terms of treatment, an effective amount is an amount that is sufficient to palliate, ameliorate, stabilize, reverse or slow the progression of the disease, or otherwise reduce the pathological consequences of the disease. The effective amount is generally determined by the physician on a case-by-case basis and is within the skill of one in the art. Several factors are typically taken into account when determining an appropriate dosage to achieve an effective amount. These factors include age, sex and weight of the subject, the condition being treated, the severity of the condition and the form and effective concentration of the immunoresponsive cells administered.

[0196] In one non-limiting example, following administration of one or more of the presently disclosed transduced cells, peripheral blood of the subject is collected and hemoglobin levels is measured. A therapeutically relevant level of hemoglobin is produced following administration of one or more of the presently disclosed transduced cells. Therapeutically relevant level of hemoglobin is a level of hemoglobin that is sufficient (1) to improve or correct anemia, (2) to restore the ability of the subject to produce red blood cells containing normal hemoglobin, (3) to correct ineffective erythropoiesis in the subject, (4) to correct extra-medullary hematopoiesis (e.g., splenic and hepatic extra-medullary hematopoiesis), and/or (5) to reduce iron accumulation, e.g., in peripheral tissues and organs. Therapeutically relevant level of hemoglobin can be at least about 7 g/dL Hb, at least about 7.5 g/dL Hb, at least about 8 g/dL Hb, at least about 8.5 g/dL Hb, at least about 9 g/dL Hb, at least about 9.5 g/dL Hb, at least about 10 g/dL Hb, at least about 10.5 g/dL Hb, at least about 11 g/dL Hb, at least about 11.5 g/dL Hb, at least about 12 g/dL Hb, at least about 12.5 g/dL Hb, at least about 13 g/dL Hb, at least about 13.5 g/dL Hb, at least about 14 g/dL Hb, at least about 14.5 g/dL Hb, or at least about 15 g/dL Hb. Additionally or alternatively, therapeutically relevant level of hemoglobin can be from about 7 g/dL Hb to about 7.5 g/dL Hb, from about 7.5 g/dL Hb to about 8 g/dL Hb, from about 8 g/dL Hb to about 8.5 g/dL Hb, from about 8.5 g/dL Hb to about 9 g/dL Hb, from about 9 g/dL Hb to about 9.5 g/dL Hb, from about 9.5 g/dL Hb to about 10 g/dL Hb, from about 10 g/dL Hb to about 10.5 g/dL Hb, from about 10.5 g/dL Hb to about 11 g/dL Hb, from about 11 g/dL Hb to about 11.5 g/dL Hb, from about 11.5 g/dL Hb to about 12 g/dL Hb, from about 12 g/dL Hb to about 12.5 g/dL Hb, from about 12.5 g/dL Hb to about 13 g/dL Hb, from about 13 g/dL Hb to about 13.5 g/dL Hb, from about 13.5 g/dL Hb to about 14 g/dL Hb, from about 14 g/dL Hb to about 14.5 g/dL Hb, from about 14.5 g/dL Hb to about 15 g/dL Hb, from about 7 g/dL Hb to about 8 g/dL Hb, from about 8 g/dL Hb to about 9 g/dL Hb, from about 9 g/dL Hb to about 10 g/dL Hb, from about 10 g/dL Hb to about 11 g/dL Hb, from about 11 g/dL Hb to about 12 g/dL Hb, from about 12 g/dL Hb to about 13 g/dL Hb, from about 13 g/dL Hb to about 14 g/dL Hb, from about 14 g/dL Hb to about 15 g/dL Hb, from about 7 g/dL Hb to about 9 g/dL Hb, from about 9 g/dL Hb to about 11 g/dL Hb, from about 11 g/dL Hb to about 13 g/dL Hb, or from about 13 g/dL Hb to about 15 g/dL Hb. In certain embodiments, the therapeutically relevant level of hemoglobin is maintained in the subject for at least about 6 months, for at least about 12 months (or 1 year), for at least about 24 months (or 2 years). In certain embodiments, the therapeutically relevant level of hemoglobin is maintained in the subject for up to about 6 months, for up to about 12 months (or 1 year), for up to about 24 months (or 2 years). In certain embodiments, the therapeutically relevant level of hemoglobin is maintained in the subject for about 6 months, for about 12 months (or 1 year), for about 24 months (or 2 years). In certain embodiments, the therapeutically relevant level of hemoglobin is maintained in the subject for from about 6 months to about 12 months (e.g., from about 6 months to about 8 months, from about 8 months to about 10 months, from about 10 months to about 12 months), from about 12 months to about 18 months (e.g., from about 12 months to about 14 months, from about 14 months to about 16 months, or from about 16 months to about 18 months), or from about 18 months to about 24 months (e.g., from about 18 months to about 20 months, from about 20 months to about 22 months, or from about 22 months to about 24 months).

[0197] In certain embodiments, the method comprises administering one or more cells transduced with a recombinant vector comprising a presently disclosed expression cassette as described above. The vector copy number of the recombinant vector in the cells that provide for the therapeutically relevant level of hemoglobin (e.g., 9-10 g/dL) in the subject is from about 0.5 to about 2, from about 0.5 to about 1, or from about 1 to about 2 vector copy number per cell. In certain embodiments, the vector copy number of the presently disclosed vector is about 0.5, about 0.6, about 0.7, about 0.8, about 0.9, about 1.0, about 1.1, about 1.2, about 1.3, about 1.4, about 1.5, about 1.6, about 1.7, about 1.8, about 1.9, or about 2.0 vector copy number per cell.

[0198] In certain embodiments, the subject lacks a human leukocyte antigen (HLA)-matched donor. In certain embodiments, the transduced cell is from the same subject. In one embodiment, the transduced cell is from bone marrow of the same subject. Thus, administration of the transduced cells do not incur the risk of graft-versus host disease in the subject. The method does not require immune suppression to prevent graft rejection, e.g., the method does not comprise administering an immunosuppressive agent to the subject.

[0199] The present disclosed subject matter also provides a method of increasing the proportion of red blood cells or erythrocytes compared to white blood cells or leukocytes in a subject. In various non-limiting embodiments, the method comprises administering an effective amount of a presently disclosed transduced cell or a population of the presently disclosed transduced cells (e.g., HSCs, embryonic stem cells, or iPSCs) to the subject, wherein the proportion of red blood cell progeny cells of the hematopoietic stem cells are increased compared to white blood cell progeny cells of the hematopoietic stem cells in the subject.

[0200] Without wishing to be bound to any particular theory, an important advantage provided by the expression cassette, vectors and other delivery systems (e.g., nucleases and CRISPR-Cas systems), compositions, and methods of the presently disclosed subject is the high efficacy of globin gene therapy that can be achieved by administering populations of cells comprising lower percentages of transduced cells compared to existing methods. This provides important safety advantages associated with reduced chances of deleterious mutation, transformation, or oncogene activation of cellular genes in transduced cells. The transduced cells can be administered as part of a bone marrow or cord blood transplant in an individual that has or has not undergone bone marrow ablative therapy.

[0201] One consideration concerning the therapeutic use of the presently disclosed cells transduced with the expression cassette described herein ("transduced cells") is the quantity of cells necessary to achieve an optimal effect. The quantity of transduced cells to be administered will vary for the subject being treated. In one embodiment, from about 1×10^4 to about 1×10^5 cells/kg, from about 1×10^5 to about 1×10^6 cells/kg, from about 1×10^6 to about 1×10^7 cells/kg, from about 1×10^7 to about 1×10^8 cells/kg, from about 1×10^8 to about 1×10^9 cells/kg, or from about 1×10^9 to about 1×10^{10} cells/kg of the presently disclosed transduced cells are administered to a subject. More effective cells may be administered in even smaller numbers. In some embodiments, at least about 1×10^8 cells/kg, at least about 2×10^8 cells/kg, at least about 3×10^8 cells/kg, at least about 4×10^8 cells/kg, or at least about 5×10^8 cells/kg of the presently disclosed transduced cells are administered to a subject. The precise determination of what would be considered an effective dose may be based on factors individual to each subject, including their size, age, sex, weight, and condition of the particular subject. Dosages can be readily ascertained by those skilled in the art from this disclosure and the knowledge in the art.

[0202] In various embodiments, the expression cassettes, vectors and other delivery systems (nucleases and CRISPR-Cas systems), compositions, and methods of the presently disclosed subject matter offer improved methods of gene therapy using *ex vivo* gene therapy and autologous transplantation. Transplantation of cells transduced with the expression cassette or into subjects having a hemoglobinopathy results in long-term correction of the disease.

[0203] One or more presently disclosed transduced cells can be administered by any methods known in the art, including, but not limited to, parenteral administration (e.g., intramuscular administration, intravenous administration, subcutaneous administration, or intraperitoneal administration), spinal administration, and epidermal administration. In one non-limiting embodiment, one or more transduced cells are delivered to a subject intravenously. One or more presently disclosed transduced cells can be administered by injection, infusion, or implantation. In one non-limiting embodiment, one or more transduced cells are administered by injection. In another non-limiting embodiment, one or more transduced cells are administered by intravenous injection.

[0204] The subjects can have an advanced form of disease, in which case the treatment objective can include mitigation or reversal of disease progression, and/or amelioration of side effects. The subjects can have a history of the condition, for which they have already been treated, in which case the therapeutic objective will typically include a decrease or delay in the risk of recurrence.

VII. Kits

[0205] The presently disclosed subject matter provides kits for the treatment or prevention of a hemoglobinopathy. In one embodiment, the kit comprises a therapeutic or prophylactic composition containing an effective amount of a cell transduced with the presently disclosed expression cassette in unit dosage form. In one non-limiting embodiment, the kit comprises one or more expression cassettes disclosed herein. In certain embodiments, the kit comprises one or more vectors comprising an expression cassette disclosed herein. In some embodiments, the kit comprises a sterile container, which can be a box, an ampule, a bottle, a vial, a tube, a bag, a pouch, a blister-pack, or other suitable container forms known in the art. Such containers can be made of plastic, glass, laminated paper, metal foil, or other materials suitable for holding medicaments.

[0206] If desired, the transduced cell is provided together with instructions for administering the cell to a subject having or at risk of developing a hemoglobinopathy. The instructions will generally include information about the use of the composition for the treatment or prevention of a hemoglobinopathy. In other embodiments, the instructions include at least one of the following: description of the therapeutic agent; dosage schedule and administration for treatment or prevention of a hemoglobinopathy or symptoms thereof; precautions; warnings; indications; counter-indications; over-dosage information; adverse reactions; animal pharmacology; clinical studies; and/or references. Alternatively or additionally, the kit can include instructions for transducing a cell with the one or more expression cassettes and/or vectors comprising such expression cassettes. The instructions may be printed directly on the container (when present), or as a label applied to the container, or as a separate sheet, pamphlet, card, or folder supplied in or with the container.

EXAMPLES

[0207] The practice of the presently disclosed subject matter employs, unless otherwise indicated, conventional tech-

niques of molecular biology (including recombinant techniques), microbiology, cell biology, biochemistry and immunology, which are well within the purview of the skilled artisan. Such techniques are explained fully in the literature, such as, "Molecular Cloning: A Laboratory Manual", second edition (Sambrook, 1989); "Oligonucleotide Synthesis" (Gait, 1984); "Animal Cell Culture" (Freshney, 1987); "Methods in Enzymology" "Handbook of Experimental Immunology" (Weir, 1996); "Gene Transfer Vectors for Mammalian Cells" (Miller and Calos, 1987); "Current Protocols in Molecular Biology" (Ausubel, 1987); "PCR: The Polymerase Chain Reaction", (Mullis, 1994); "Current Protocols in Immunology" (Coligan, 1991). These techniques are applicable to the production of the polynucleotides and polypeptides of the presently disclosed subject matter, and, as such, may be considered in making and practicing the presently disclosed subject matter. Particularly useful techniques for particular embodiments will be discussed in the sections that follow.

[0208] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the expression cassettes, vectors, delivery systems, and therapeutic methods of the presently disclosed subject matter, and are not intended to limit the scope of what the inventors regard as their invention.

Example 1: Discovery Of Novel Insulators

[0209] The problems created by insertional mutagenesis of viral vectors are widely known (Nienhuis (2013), Baum *et al.* (2006), Nienhuis *et al.* (2006)) as is the evidence that the risks of genotoxicity can be reduced by the use of chromatin insulators (Arumugam *et al.* (2007), Emery (2011), Evans-Galea *et al.* (2007), Rivella *et al.* (2000), Emery *et al.* (2000), Emery *et al.* (2002), Yannaki *et al.* (2002), Hino *et al.* (2004), Ramezani *et al.* (2003), Ramezani *et al.* (2008)). Approaches allowing the efficient identification of enhancer blocking insulators in the human genome have been developed. These new insulators are short, on the average 150bp, and they do not affect adversely the titers of viral vectors and they are several times more powerful than the insulator cHS4. Genomic approaches were used to discover the most powerful enhancer blocker and barrier insulators of the human genome. For gene therapy of the hemoglobinopathies, powerful enhancers are required to achieve therapeutic levels of globin gene expression. Powerful insulators may thus provide one means to protect the genomic environment from the powerful enhancers of the integrating vectors.

[0210] Several studies have demonstrated the ability of the cHS4 insulator to reduce position-effect silencing of gammaretroviral vectors (Evans-Galea *et al.* (2007), Rivella *et al.* (2000), Emery *et al.* (2000), Emery *et al.* (2002), Yannaki *et al.* (2002), Hino *et al.* (2004), Ramezani *et al.* (2006), Yao *et al.* (2003), Nishino *et al.* (2006), Aker *et al.* (2007), Li and Emery (2008)), and lentiviral vectors (Evans-Galea *et al.* (2007), Ramezani *et al.* (2003), Puthenveetil *et al.* (2004), Arumugam *et al.* (2007), Bank *et al.* (2005), Aker *et al.* (2007), Ma *et al.* (2003), Chang *et al.* (2005), Pluta *et al.* (2005)). Those studies that were appropriately designed demonstrated that inclusion of the 1.2 kb version of the cHS4 insulator increased the likelihood and/or consistency of vector transgene expression in at least some settings (Arumugam *et al.* (2007), Evans-Galea *et al.* (2007), Emery *et al.* (2002), Yannaki *et al.* (2002), Hino *et al.* (2004), Ramezani *et al.* (2006), Aker *et al.* (2007), Li and Emery (2008), Pluta *et al.* (2005), Jakobsson *et al.* (2004)). Nevertheless, the degree of protection afforded by the cHS4 insulator is far from complete. In addition, the inclusion of the 1.2 Kb cHS4 can adversely affect vector titers while the smallest cHS4 core has been proven ineffective (Aker *et al.* (2007), Jakobsson *et al.* (2004)).

[0211] Effects on genotoxicity were tested using an *in vivo* assay based on quantitation of tumor formation in mice. Vectors insulated by insulator A1 decreased tumor formation induced by random vector integration in hematopoietic chimeras compared to mice that received uninsulated or cHS4-insulated controls.

[0212] To assess effects on vector titers, insulator A1 was introduced into the double-copy region of a third-generation lentiviral vector expressing GFP from a constitutive package promoter, and the viral titers and GFP expression were measured. Insulator A1 did not affect adversely vector GFP expression.

[0213] In the *in vivo* genotoxicity assay, a cell line transduced with gammaretroviral vectors produced tumors after transplantation in mice and allowed quantitation of genotoxic effects by measuring rates of tumor free survival. Effects of an insulator on genotoxicity were quantitated by the number of tumors formed in the mice and the rates of tumor free survival. Insulator A1 was inserted in the proximal portion of the 3' LTR, from which it is copied into the 5' LTR during reverse transcription and vector integration. The resulting topology places copies of the insulator between the genomic regions located 5' and 3' of the integrated provirus and enhancer activity from the 5' viral LTR and internal Pgk promoter, but does not contain the enhancer in the 3' LTR. This can decrease genotoxicity thus resulting in decreased tumor formation and increased survival of the animals. Gamma-retroviral reporter vectors flanked with insulator A1 or control regions were used to transduce the growth factor-dependent cell line 32D, and 10 independent sub-pools for each vector were transplanted into syngeneic C3H/HeJ mice. All 10 mice transplanted with mock-transduced cells remained free of 32D cell- derived tumors, while nearly all mice transplanted with 32D cells transduced with vectors containing no inserts or a 790 bp neutral spacer developed tumors within a median of 16 weeks (Figure 5B). Flanking this vector with the cHS4 insulator delayed the onset of tumor formation by several weeks, and reduced the frequency of animals that developed tumors to 6 of 10. In contrast, only two of 10 animals developed tumors following transplantation with 32D cells transduced with the vector flanked with insulator A1 (Figure 5B). The frequency of animals with tumors and the

number of vector transduction events in the original sub-pools suggested that flanking the vector with insulator A1 reduced the overall rate of tumor formation 12-fold, from 46.9 tumors per 10⁵ provirus to 3.9 tumors per 10⁵ provirus (Figure 5C). In comparison, the cHS4 insulator reduced the overall rate of tumor formation 2.8-fold (to 16.9 tumors per 10⁵ provirus), while the neutral spacer had no statistically discernable effect on the rate of tumor formation. These results indicate that the discovered enhancer blocking insulators can decrease substantially the risks of insertional mutagenesis and genotoxicity.

Example 2: Characterization Of Globin Vectors Comprising At Least One Insulator.

[0214] A presently disclosed expression cassette (designated as "Expression Cassette 1"; as shown in Figure 1), which comprises insulator A1, and a human β^A -globin gene encoding a threonine to glutamine mutation at codon 87 (β^A -T87Q) operably linked to a β -globin LCR region comprising a HS2 region having the nucleotide sequence set forth in SEQ ID NO:9, a HS3 region having the nucleotide sequence set forth in SEQ ID NO:5, and a HS4 region having the nucleotide sequence set forth in SEQ ID NO:7, was generated. The rationale for using the variant β chain (β^A) is to facilitate the detection of the vector-encoded β -globin gene, distinguishing it from endogenous or transfused beta chains. The glutamine (GLN) residue at position 87 in the γ -globin chain augments the anti-sickling activity of the gamma chain relative to the β chain, while preserving adult oxygen-binding characteristics of the β chain (Nagel *et al.* (1979)). In Vector 1, a point mutation altering codon 87 (β^A -T87Q, or β 87) replaces the normal threonine with glutamine and augments anti-sickling activity of the vector-encoded β chain. This β 87 chain has been safely used in a patient with HbE-thalassemia (Cavazzana-Calvo *et al.* (2010)).

[0215] Expression cassette 1 was incorporated or introduced to a lentivirus vector (designated as "Vector 1"). Vector 1 was introduced in bone marrow cells of C57BL/6-Hbb th3/+ mice and transplanted to syngeneic lethally irradiated recipients as previously described (May *et al.* (2000), May *et al.* (2002), Lisowski *et al.* (2007)). The vector titer of VI was comparable to that of a lentivirus vector comprising an expression cassette lacking insulator A1. The β -globin expression of Vector 1 was compared to that of a lentivirus vector (designated as "Vector 2") comprising an expression cassette that lacks an insulator and comprises a wild human β -globin gene operably linked to a β -globin LCR region comprising a HS2 region having the nucleotide sequence set forth in SEQ ID NO:9, a HS3 region having the nucleotide sequence set forth in SEQ ID NO:5, and a HS4 region having the nucleotide sequence set forth in SEQ ID NO:6. In comparison to Vector 2, β -globin expression of Vector 1 normalized to vector copy was equivalent or slightly increased, suggesting an added benefit for *in vivo* expression provided by the flanking barrier elements, as shown in Figure 6.

Example 3: Evaluation of Enhancer Activity in Non-erythroid K562 Cells

[0216] The enhancer activity of HS2 was evaluated in Non-erythroid K562 Cells. As shown in Figure 7, GFP expression in K562 cells transduced with vectors driven by a minimal promoter linked to no enhancer ("Empty", HS2, HS3-4, HS2-3-4 or the runx1 enhancer used as positive control ("RUNX1"). Background expression was on the order or 0.01% ("empty"), but increased over 10-fold with HS2-3-4 ("Lcr9", 0.17%). This enhancement was mostly due to HS2 (0.15%) but not HS3-4 (0.05%). All cell lines were comparably transduced (mean vector copy number 2.5). The results support that HS2 but not HS3-HS4 may pose an oncogenic risk in non-erythroid hematopoietic stem and progenitor cells.

Example 4: Novel erythroid-specific enhancers

[0217] As shown in Figures 8 and 9, five erythroid-specific enhancers were substituted for HS2: ALAS Intron 1, ALAS Intron 8, BLVRB, PPOX, and Spectrin-alpha. The inventors have shown that all these five enhancers are powerful enhancers, and lack enhancer activity in non-erythroid tissues, and do not reduce the vector titer.

Example 5: Increasing globin lentiviral vector production through 3' LTR modifications

[0218] An essential feature of therapeutic globin vectors is to achieve a high titer, sufficient for effective transduction of patient cells. By virtue of their large cargo, comprising a gene, promoter, enhancers and/or LCR elements, globin lentiviral vectors inherently have low titer, complicating their manufacture and limiting their clinical use. This problem is further compounded by the incorporation of additional genomic elements such as an insulator, which further increase the size of the vector.

[0219] The inventors explored different modifications of the 3' long terminal repeat (LTR) of globin vectors to increase the titer of globin vectors. Over 62 variations were evaluated, numbered 1 through 62, modeled on a lentivirus vector comprising a human β -globin gene operably linked to a β -globin LCR region comprising a HS2 region having the nucleotide sequence set forth in SEQ ID NO:9, a HS3 region having the nucleotide sequence set forth in SEQ ID NO:5, and a HS4 region having the nucleotide sequence set forth in SEQ ID NO:7. In other words, all of Vectors #1 through Vector 62

comprise a β -globin LCR region comprising a HS2 region having the nucleotide sequence set forth in SEQ ID NO:9, a HS3 region having the nucleotide sequence set forth in SEQ ID NO:5, and a HS4 region having the nucleotide sequence set forth in SEQ ID NO:7. Vector #18 served as a baseline, comprising a standard U3 deletion in the 3'LTR. Vector #1 (not depicted) comprised a full, i.e., wild-type LTR, which cannot be used clinically. Modifications to the 3'LTR are depicted in Figures 10A and 10B, and their titers shown in Figures 11 and 12 (the Y axis shows the vector copy number of vector stocks manufactured and tested under strictly identical conditions). Titrations were measured in triple replicas, performed in parallel by two operators, and repeated in multiple experiments.

[0220] As shown in Figures 11 and 12, Vector #55 repeatedly showed a higher titer. This vector comprises a Woodchuck hepatitis post-regulatory element (WPRE) and a bovine growth hormone polyadenylation signal 3' to the R region in the 3' LTR. The WPRE element is therefore not transferred to the transduced cells.

[0221] The incorporation of these elements for enhancing the production of globin lentiviral vectors is essential to yield higher titers and hence for the clinical usefulness of the vectors described in this application.

REFERENCES

[0222]

1. Weatherall, D.J. & Clegg, J.B. The Thalassemia Syndrome. Blackwell Scientific Oxford (1981).
2. Stamatoyannopoulos, G., Nienhuis, A.W., Majerus, P. & Varmus, H. The Molecular Basis of Blood Diseases. WB Saunders, Philadelphia (1994).
3. Weatherall, D.J. Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. *Nat Rev Genet* 2, 245-255. (2001).
4. Steinberg, M.H., Forget, B.G., Higgs, D.R. & Nagel, R.L. Molecular Mechanism of β Thalassemia; Bernard G. Forget, (Cambridge University Press, Cambridge, UK, 2001).
5. Cooley, T.B. & Lee, P. A series of cases of splenomegaly in children with anemia and peculiar bone changes. *Trans. Am. Pediatr. Soc.* 37, 29 (1925).
6. Giardina, P.J. & Grady, R.W. Chelation therapy in beta-thalassemia: an optimistic update. *Semin Hematol* 38, 360-366. (2001).
7. Giardini, C. & Lucarelli, G. Bone marrow transplantation in the treatment of thalassemia. *Current opinion in hematology* 1, 170-176. (1994).
8. Boulad, F., Giardina, P., Gillio, A., Kernan, N., Small, T., Brochstein, J., Van Syckle, K., George, D., Szabolcs, P. & O'Reilly, R.J. Bone marrow transplantation for homozygous beta-thalassemia. The Memorial Sloan-Kettering Cancer Center experience. *Ann N Y Acad Sci* 850, 498-502. (1998).
9. Lucarelli, G., Clift, R.A., Galimberti, M., Angelucci, E., Giardini, C., Baronciani, D., Polchi, P., Andreani, M., Gaziev, D., Erer, B., Ciaroni, A., D'Adamo, F., Albertini, F. & Muretto, P. Bone marrow transplantation in adult thalassemic patients. *Blood* 93, 1164-1167. (1999).
10. Tisdale, J. & Sadelain, M. Toward gene therapy for disorders of globin synthesis. *Semin Hematol* 38, 382-392 (2001).
11. Pauling, L., Itano, H.A., Singer, S.J. & Wells, I.C. Sickle cell anemia, a molecular disease. *Science* 110, 543-546 (1949).
12. Swank, R.A. & Stamatoyannopoulos, G. Fetal gene reactivation. *Curr Opin Genet Dev* 8, 366-370 (1998).
13. Platt, O.S., Orkin, S.H., Dover, G., Beardsley, G.P., Miller, B. & Nathan, D.G. Hydroxyurea enhances fetal hemoglobin production in sickle cell anemia. *J Clin Invest* 74, 652-656. (1984).
14. Charache, S., Dover, G.J., Moore, R.D., Eckert, S., Ballas, S.K., Koshy, M., Milner, P.F., Orringer, E.P., Phillips, G., Jr., Platt, O.S. & et al. Hydroxyurea: effects on hemoglobin F production in patients with sickle cell anemia. *Blood* 79, 2555-2565. (1992).
15. Atweh, G.F. & Loukopoulos, D. Pharmacological induction of fetal hemoglobin in sickle cell disease and beta-thalassemia. *Semin Hematol* 38, 367-373. (2001).
16. Perrine, S.P., Castaneda, S.A., Boosalis, M.S., White, G.L., Jones, B.M. & Bohacek, R. Induction of fetal globin in beta-thalassemia: Cellular obstacles and molecular progress. *Ann N Y Acad Sci* 1054, 257-265 (2005).
17. Stamatoyannopoulos, G. Prospects for developing a molecular cure for thalassemia. *Hematology* 10 Suppl 1, 255-257 (2005).
18. Vermynen, C., Cornu, G., Ferster, A., Brichard, B., Ninane, J., Ferrant, A., Zenebergh, A., Maes, P., Dhooge, C., Benoit, Y., Beguin, Y., Dresse, M.F. & Sariban, E. Haematopoietic stem cell transplantation for sickle cell anaemia: the first 50 patients transplanted in Belgium. *Bone Marrow Transplant* 22, 1-6 (1998).
19. Luzzatto, L. & Goodfellow, P. Sickle cell anaemia. A simple disease with no cure. *Nature* 337, 17-18 (1989).
20. Sadelain, M. Genetic treatment of the haemoglobinopathies: recombinations and new combinations. *Br J Haematol* 98, 247-253 (1997).

21. Sadelain, M., Boulad, F., Galanello, R., Giardina, P., Locatelli, F., Maggio, A., Rivella, S., Riviere, I. & Tisdale, J. Therapeutic options for patients with severe beta-thalassemia: the need for globin gene therapy. *Hum Gene Ther* 18, 1-9 (2007).
22. Borgna-Pignatti, C., Rugolotto, S., De Stefano, P., Zhao, H., Cappellini, M.D., Del Vecchio, G.C., Romeo, M.A., Forni, G.L., Gamberini, M.R., Ghilardi, R., Piga, A. & Cnaan, A. Survival and complications in patients with thalassemia major treated with transfusion and deferoxamine. *Haematologica* 89, 1187-1193 (2004).
23. Telfer, P.T., Warburton, F., Christou, S., Hadjigavriel, M., Sitarou, M., Kolnagou, A. & Angastiniotis, M. Improved survival in thalassemia major patients on switching from desferrioxamine to combined chelation therapy with desferrioxamine and deferiprone. *Haematologica* 94, 1777-1778 (2009).
24. Ladis, V., Chouliaras, G., Berdoukas, V., Chatziliami, A., Fragodimitri, C., Karabatsos, F., Youssef, J., Kattamis, A. & Karagiorga-Lagana, M. Survival in a large cohort of Greek patients with transfusion-dependent beta thalassaemia and mortality ratios compared to the general population. *European journal of haematology* 86, 332-338 (2011).
25. Mancuso, A., Sciarrino, E., Renda, M.C. & Maggio, A. A prospective study of hepatocellular carcinoma incidence in thalassemia. *Hemoglobin* 30, 119-124 (2006).
26. Persons, D.A. & Tisdale, J.F. Gene therapy for the hemoglobin disorders. *Semin Hematol* 41, 279-286 (2004).
27. Sadelain, M. Recent advances in globin gene transfer for the treatment of beta-thalassemia and sickle cell anemia. *Current opinion in hematology* 13, 142-148 (2006).
28. May, C., Rivella, S., Callegari, J., Heller, G., Gaensler, K.M., Luzzatto, L. & Sadelain, M. Therapeutic haemoglobin synthesis in beta-thalassaemic mice expressing lentivirus-encoded human beta-globin. *Nature* 406, 82-86 (2000).
29. May, C., Rivella, S., Chadburn, A. & Sadelain, M. Successful treatment of murine beta-thalassemia intermedia by transfer of the human beta-globin gene. *Blood* 99, 1902-1908 (2002).
30. Rivella, S., May, C., Chadburn, A., Riviere, I. & Sadelain, M. A novel murine model of Cooley anemia and its rescue by lentiviral-mediated human beta-globin gene transfer. *Blood* 101, 2932-2939 (2003).
31. Sadelain, M., Boulad, F., Lisowski, L., Moi, P. & Riviere, I. Stem cell engineering for the treatment of severe hemoglobinopathies. *Curr Mol Med* 8, 690-697 (2008).
32. Bank, A., Dorazio, R. & Leboulch, P. A phase I/II clinical trial of beta-globin gene therapy for beta-thalassemia. *Ann N Y Acad Sci* 1054, 308-316 (2005).
33. Cavazzana-Calvo, M., Payen, E., Negre, O., Wang, G., Hehir, K., Fusil, F., Down, J., Denaro, M., Brady, T., Westerman, K., Cavallese, R., Gillet-LeGrand, B., Caccavelli, L., Sgarra, R., Maouche-Chretien, L., Bernaudin, F., Girot, R., Dorazio, R., Mulder, G.J., Polack, A., Bank, A., Soulier, J., Larghero, J., Kabbara, N., Dalle, B., Gourmel, B., Socie, G., Chretien, S., Cartier, N., Aubourg, P., Fischer, A., Cornetta, K., Galacteros, F., Beuzard, Y., Gluckman, E., Bushman, F., Hacein-Bey-Abina, S. & Leboulch, P. Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia. *Nature* 467, 318-322 (2010).
34. Braun, C.J., Boztug, K., Paruzynski, A., Witzel, M., Schwarzer, A., Rothe, M., Modlich, U., Beier, R., Gohring, G., Steinemann, D., Fronza, R., Ball, C.R., Haemmerle, R., Naundorf, S., Kuhlcke, K., Rose, M., Fraser, C., Mathias, L., Ferrari, R., Abboud, M.R., Al-Herz, W., Kondratenko, I., Marodi, L., Glimm, H., Schlegelberger, B., Schambach, A., Albert, M.H., Schmidt, M., von Kalle, C. & Klein, C. Gene therapy for Wiskott-Aldrich syndrome--long-term efficacy and genotoxicity. *Sci Transl Med* 6, 227ra233 (2014).
35. Chang, A.H. & Sadelain, M. The genetic engineering of hematopoietic stem cells: the rise of lentiviral vectors, the conundrum of the Itr, and the promise of lineage-restricted vectors. *Mol Ther* 15, 445-456 (2007).
36. Pawliuk, R., Westerman, K.A., Fabry, M.E., Payen, E., Tighe, R., Bouhassira, E.E., Acharya, S.A., Ellis, J., London, I.M., Eaves, C.J., Humphries, R.K., Beuzard, Y., Nagel, R.L. & Leboulch, P. Correction of sickle cell disease in transgenic mouse models by gene therapy. *Science* 294, 2368-2371 (2001).
37. Emery, D.W., Chen, H., Li, Q. & Stamatoyannopoulos, G. Development of a condensed locus control region cassette and testing in retrovirus vectors for A gamma-globin. *Blood Cells Mol Dis* 24, 322-339 (1998).
38. Miccio, A., Cesari, R., Lotti, F., Rossi, C., Sanvito, F., Ponzoni, M., Routledge, S.J., Chow, C.M., Antoniou, M.N. & Ferrari, G. In vivo selection of genetically modified erythroblastic progenitors leads to long-term correction of beta-thalassemia. *Proc Natl Acad Sci USA* 105, 10547-10552 (2008).
39. Sadelain, M., Wang, C.H., Antoniou, M., Grosveld, F. & Mulligan, R.C. Generation of a high-titer retroviral vector capable of expressing high levels of the human beta-globin gene. *Proc Natl Acad Sci USA* 92, 6728-6732 (1995).
40. Samakoglu, S., Lisowski, L., Budak-Alpdogan, T., Usachenko, Y., Acuto, S., Di Marzo, R., Maggio, A., Zhu, P., Tisdale, J.F., Riviere, I. & Sadelain, M. A genetic strategy to treat sickle cell anemia by coregulating globin transgene expression and RNA interference. *Nat Biotechnol* 24, 89-94 (2006).
41. Pestina, T.I., Hargrove, P.W., Jay, D., Gray, J.T., Boyd, K.M. & Persons, D.A. Correction of murine sickle cell disease using gamma-globin lentiviral vectors to mediate high-level expression of fetal hemoglobin. *Mol Ther* 17, 245-252 (2009).
42. Hanawa, H., Yamamoto, M., Zhao, H., Shimada, T. & Persons, D.A. Optimized lentiviral vector design improves

titer and transgene expression of vectors containing the chicken beta-globin locus HS4 insulator element. *Mol Ther* 17, 667-674 (2009).

43. Arumugam, P.I., Scholes, J., Perelman, N., Xia, P., Yee, J.K. & Malik, P. Improved human beta-globin expression from self-inactivating lentiviral vectors carrying the chicken hypersensitive site-4 (cHS4) insulator element. *Mol Ther* 15, 1863-1871 (2007).

44. Fraser, P., Pruzina, S., Antoniou, M. & Grosveld, F. Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes. *Genes & development* 7, 106-113 (1993).

45. Navas, P.A., Peterson, K.R., Li, Q., Skarpidi, E., Rohde, A., Shaw, S.E., Clegg, C.H., Asano, H. & Stamatoyannopoulos, G. Developmental specificity of the interaction between the locus control region and embryonic or fetal globin genes in transgenic mice with an HS3 core deletion. *Molecular and cellular biology* 18, 4188-4196 (1998).

46. Li, Q. & Stamatoyannopoulos, G. Hypersensitive site 5 of the human beta locus control region functions as a chromatin insulator. *Blood* 84, 1399-1401 (1994).

47. Li, Q., Zhang, M., Han, H., Rohde, A. & Stamatoyannopoulos, G. Evidence that DNase I hypersensitive site 5 of the human beta-globin locus control region functions as a chromosomal insulator in transgenic mice. *Nucleic Acids Res* 30, 2484-2491 (2002).

48. Puthenveetil, G., Scholes, J., Carbonell, D., Qureshi, N., Xia, P., Zeng, L., Li, S., Yu, Y., Hiti, A.L., Yee, J.K. & Malik, P. Successful correction of the human beta-thalassemia major phenotype using a lentiviral vector. *Blood* 104, 3445-3453 (2004).

49. Wilber, A., Nienhuis, A.W. & Persons, D.A. Transcriptional regulation of fetal to adult hemoglobin switching: new therapeutic opportunities. *Blood* 117, 3945-3953 (2011).

50. Arumugam, P.I., Higashimoto, T., Urbinati, F., Modlich, U., Nestheide, S., Xia, P., Fox, C., Corsinotti, A., Baum, C. & Malik, P. Genotoxic potential of lineage-specific lentivirus vectors carrying the beta-globin locus control region. *Mol Ther* 17, 1929-1937 (2009).

51. Chang, K.H., Fang, X., Wang, H., Huang, A., Cao, H., Yang, Y., Bonig, H., Stamatoyannopoulos, J.A. & Papayannopoulou, T. Epigenetic modifications and chromosome conformations of the beta globin locus throughout development. *Stem cell reviews* 9, 397-407 (2013).

52. Papayannopoulou, T., Priestley, G.V., Rohde, A., Peterson, K.R. & Nakamoto, B. Hemopoietic lineage commitment decisions: in vivo evidence from a transgenic mouse model harboring micro LCR-beta_{pro}-LacZ as a transgene. *Blood* 95, 1274-1282 (2000).

53. Nienhuis, A.W. Development of gene therapy for blood disorders: an update. *Blood* 122, 1556-1564 (2013).

54. Baum, C., Kustikova, O., Modlich, U., Li, Z. & Fehse, B. Mutagenesis and oncogenesis by chromosomal insertion of gene transfer vectors. *Hum Gene Ther* 17, 253-263 (2006).

55. Nienhuis, A.W., Dunbar, C.E. & Sorrentino, B.P. Genotoxicity of retroviral integration in hematopoietic cells. *Mol Ther* 13, 1031-1049 (2006).

56. Emery, D.W. The use of chromatin insulators to improve the expression and safety of integrating gene transfer vectors. *Hum Gene Ther* 22, 761-774 (2011).

57. Evans-Galea, M.V., Wielgosz, M.M., Hanawa, H., Srivastava, D.K. & Nienhuis, A.W. Suppression of clonal dominance in cultured human lymphoid cells by addition of the cHS4 insulator to a lentiviral vector. *Mol Ther* 15, 801-809 (2007).

58. Rivella, S., Callegari, J.A., May, C., Tan, C.W. & Sadelain, M. The cHS4 insulator increases the probability of retroviral expression at random chromosomal integration sites. *J Virol* 74, 4679-4687 (2000).

59. Emery, D.W., Yannaki, E., Tubb, J. & Stamatoyannopoulos, G. A chromatin insulator protects retrovirus vectors from chromosomal position effects. *Proc Natl Acad Sci USA* 97, 9150-9155 (2000).

60. Emery, D.W., Yannaki, E., Tubb, J., Nishino, T., Li, Q. & Stamatoyannopoulos, G. Development of virus vectors for gene therapy of beta chain hemoglobinopathies: flanking with a chromatin insulator reduces gamma-globin gene silencing in vivo. *Blood* 100, 2012-2019 (2002).

61. Yannaki, E., Tubb, J., Aker, M., Stamatoyannopoulos, G. & Emery, D.W. Topological constraints governing the use of the chicken HS4 chromatin insulator in oncoretrovirus vectors. *Mol Ther* 5, 589-598 (2002).

62. Hino, S., Fan, J., Taguwa, S., Akasaka, K. & Matsuoka, M. Sea urchin insulator protects lentiviral vector from silencing by maintaining active chromatin structure. *Gene Ther* 11, 819-828 (2004).

63. Ramezani, A., Hawley, T.S. & Hawley, R.G. Performance- and safety-enhanced lentiviral vectors containing the human interferon-beta scaffold attachment region and the chicken beta-globin insulator. *Blood* 101, 4717-4724 (2003).

64. Ramezani, A., Hawley, T.S. & Hawley, R.G. Combinatorial incorporation of enhancer-blocking components of the chicken beta-globin 5'HS4 and human T-cell receptor alpha/delta BEAD-1 insulators in self-inactivating retroviral vectors reduces their genotoxic potential. *Stem Cells* 26, 3257-3266 (2008).

65. Yannaki, E., Emery, D.W. & Stamatoyannopoulos, G. Gene therapy for beta-thalassaemia: the continuing chal-

lenge. *Expert reviews in molecular medicine* 12, e31 (2010).

66. Persons, D.A. The challenge of obtaining therapeutic levels of genetically modified hematopoietic stem cells in beta-thalassemia patients. *Ann N Y Acad Sci* 1202, 69-74 (2010).

67. Perumbeti, A. & Malik, P. Therapy for beta-globinopathies: a brief review and determinants for successful and safe correction. *Ann N Y Acad Sci* 1202, 36-44 (2010).

68. Johnson, K.D., Grass, J.A., Park, C., Im, H., Choi, K. & Bresnick, E.H. Highly restricted localization of RNA polymerase II within a locus control region of a tissue-specific chromatin domain. *Molecular and cellular biology* 23, 6484-6493 (2003).

69. Vieira, K.F., Levings, P.P., Hill, M.A., Crusselle, V.J., Kang, S.H., Engel, J.D. & Bungert, J. Recruitment of transcription complexes to the beta-globin gene locus in vivo and in vitro. *J Biol Chem* 279, 50350-50357 (2004).

70. Levings, P.P., Zhou, Z., Vieira, K.F., Crusselle-Davis, V.J. & Bungert, J. Recruitment of transcription complexes to the beta-globin locus control region and transcription of hypersensitive site 3 prior to erythroid differentiation of murine embryonic stem cells. *The FEBS journal* 273, 746-755 (2006).

71. Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* 421, 448-453 (2003).

72. Felsenfeld, G. Chromatin as an essential part of the transcriptional mechanism. *Nature* 355, 219-224 (1992).

73. Brownell, J.E. & Allis, C.D. Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. *Curr Opin Genet Dev* 6, 176-184 (1996).

74. Kingston, R.E. & Narlikar, G.J. ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. *Genes & development* 13, 2339-2352 (1999).

75. Tsukiyama, T. & Wu, C. Chromatin remodeling and transcription. *Curr Opin Genet Dev* 7, 182-191 (1997).

76. Wolffe, A.P., Wong, J. & Pruss, D. Activators and repressors: making use of chromatin to regulate transcription. *Genes to cells : devoted to molecular & cellular mechanisms* 2, 291-302 (1997).

77. Kadonaga, J.T. Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. *Cell* 92, 307-313 (1998).

78. Struhl, K. Histone acetylation and transcriptional regulatory mechanisms. *Genes & development* 12, 599-606 (1998).

79. Gross, D.S. & Garrard, W.T. Nuclease hypersensitive sites in chromatin. *Annual review of biochemistry* 57, 159-197 (1988).

80. Elgin, S.C. Anatomy of hypersensitive sites. *Nature* 309, 213-214 (1984).

81. Wu, C. The 5' ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNase I. *Nature* 286, 854-860 (1980).

82. Felsenfeld, G., Boyes, J., Chung, J., Clark, D. & Studitsky, V. Chromatin structure and gene expression. *Proc Natl Acad Sci U S A* 93, 9384-9388 (1996).

83. Burgess-Beusse, B., Farrell, C., Gaszner, M., Litt, M., Mutskov, V., Recillas-Targa, F., Simpson, M., West, A. & Felsenfeld, G. The insulation of genes from external enhancers and silencing chromatin. *Proc Natl Acad Sci U S A* 99 Suppl 4, 16433-16437 (2002).

84. Elgin, S.C. DNAase I-hypersensitive sites of chromatin. *Cell* 27, 413-415 (1981).

85. McGhee, J.D., Wood, W.I., Dolan, M., Engel, J.D. & Felsenfeld, G. A 200 base pair region at the 5' end of the chicken adult beta-globin gene is accessible to nuclease digestion. *Cell* 27, 45-55 (1981).

86. Lowrey, C.H., Bodine, D.M. & Nienhuis, A.W. Mechanism of DNase I hypersensitive site formation within the human globin locus control region. *Proc Natl Acad Sci USA* 89, 1143-1147 (1992).

87. Adams, C.C. & Workman, J.L. Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Molecular and cellular biology* 15, 1405-1421 (1995).

88. McArthur, M., Gerum, S. & Stamatoyannopoulos, G. Quantification of DNaseI-sensitivity by real-time PCR: quantitative analysis of DNaseI-hypersensitivity of the mouse beta-globin LCR. *J Mol Biol* 313, 27-34 (2001).

89. Dorschner, M.O., Hawrylycz, M., Humbert, R., Wallace, J.C., Shafer, A., Kawamoto, J., Mack, J., Hall, R., Goldy, J., Sabo, P.J., Kohli, A., Li, Q., McArthur, M. & Stamatoyannopoulos, J.A. High-throughput localization of functional elements by quantitative chromatin profiling. *Nat Methods* 1, 219-225 (2004).

90. Sabo, P.J., Kuehn, M.S., Thurman, R., Johnson, B.E., Johnson, E.M., Cao, H., Yu, M., Rosenzweig, E., Goldy, J., Haydock, A., Weaver, M., Shafer, A., Lee, K., Neri, F., Humbert, R., Singer, M.A., Richmond, T.A., Dorschner, M.O., McArthur, M., Hawrylycz, M., Green, R.D., Navas, P.A., Noble, W.S. & Stamatoyannopoulos, J.A. Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat Methods* 3, 511-518 (2006).

91. Sabo, P.J., Hawrylycz, M., Wallace, J.C., Humbert, R., Yu, M., Shafer, A., Kawamoto, J., Hall, R., Mack, J., Dorschner, M.O., McArthur, M. & Stamatoyannopoulos, J.A. Discovery of functional noncoding elements by digital analysis of chromatin structure. *Proc Natl Acad Sci USA* 101, 16837-16842 (2004).

92. Sabo, P.J., Humbert, R., Hawrylycz, M., Wallace, J.C., Dorschner, M.O., McArthur, M. & Stamatoyannopoulos, J.A. Genome-wide identification of DNaseI hypersensitive sites using active chromatin sequence libraries. *Proc Natl Acad Sci USA* 101, 4537-4542 (2004).

93. Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., Garg, K., John, S., Sandstrom, R., Bates, D., Boatman, L., Canfield, T.K., Diegel, M., Dunn, D., Ebersol, A.K., Frum, T., Giste, E., Johnson, A.K., Johnson, E.M., Kutayavin, T., Lajoie, B., Lee, B.K., Lee, K., London, D., Lotakis, D., Neph, S., Neri, F., Nguyen, E.D., Qu, H., Reynolds, A.P., Roach, V., Safi, A., Sanchez, M.E., Sanyal, A., Shafer, A., Simon, J.M., Song, L., Vong, S., Weaver, M., Yan, Y., Zhang, Z., Zhang, Z., Lenhard, B., Tewari, M., Dorschner, M.O., Hansen, R.S., Navas, P.A., Stamatoyannopoulos, G., Iyer, V.R., Lieb, J.D., Sunyaev, S.R., Akey, J.M., Sabo, P.J., Kaul, R., Furey, T.S., Dekker, J., Crawford, G.E. & Stamatoyannopoulos, J.A. The accessible chromatin landscape of the human genome. *Nature* 489, 75-82 (2012).
94. Stergachis, A.B., Neph, S., Reynolds, A., Humbert, R., Miller, B., Paige, S.L., Vernot, B., Cheng, J.B., Thurman, R.E., Sandstrom, R., Haugen, E., Heimfeld, S., Murry, C.E., Akey, J.M. & Stamatoyannopoulos, J.A. Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. *Cell* 154, 888-903 (2013).
95. Neph, S., Stergachis, A.B., Reynolds, A., Sandstrom, R., Borenstein, E. & Stamatoyannopoulos, J.A. Circuitry and dynamics of human transcription factor regulatory networks. *Cell* 150, 1274-1286 (2012).
96. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kutayavin, T., Stehling-Sun, S., Johnson, A.K., Canfield, T.K., Giste, E., Diegel, M., Bates, D., Hansen, R.S., Neph, S., Sabo, P.J., Heimfeld, S., Raubitschek, A., Ziegler, S., Cotsapas, C., Sotoodehnia, N., Glass, I., Sunyaev, S.R., Kaul, R. & Stamatoyannopoulos, J.A. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337, 1190-1195 (2012).
97. Stergachis, A.B., Haugen, E., Shafer, A., Fu, W., Vernot, B., Reynolds, A., Raubitschek, A., Ziegler, S., LeProust, E.M., Akey, J.M. & Stamatoyannopoulos, J.A. Exonic transcription factor binding directs codon choice and affects protein evolution. *Science* 342, 1367-1372 (2013).
98. Neph, S., Vierstra, J., Stergachis, A.B., Reynolds, A.P., Haugen, E., Vernot, B., Thurman, R.E., John, S., Sandstrom, R., Johnson, A.K., Maurano, M.T., Humbert, R., Rynes, E., Wang, H., Vong, S., Lee, K., Bates, D., Diegel, M., Roach, V., Dunn, D., Neri, J., Schafer, A., Hansen, R.S., Kutayavin, T., Giste, E., Weaver, M., Canfield, T., Sabo, P., Zhang, M., Balasundaram, G., Byron, R., MacCoss, M.J., Akey, J.M., Bender, M.A., Groudine, M., Kaul, R. & Stamatoyannopoulos, J.A. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* 489, 83-90 (2012).
99. Ramezani, A., Hawley, T.S. & Hawley, R.G. Stable gammaretroviral vector expression during embryonic stem cell-derived in vitro hematopoietic development. *Mol Ther* 14, 245-254 (2006).
100. Recillas-Targa, F., Pikaart, M.J., Burgess-Beusse, B., Bell, A.C., Litt, M.D., West, A.G., Gaszner, M. & Felsenfeld, G. Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc Natl Acad Sci USA* 99, 6883-6888 (2002).
101. Gaszner, M. & Felsenfeld, G. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet* 7, 703-713 (2006).
102. Wallace, J.A. & Felsenfeld, G. We gather together: insulators and genome organization. *Curr Opin Genet Dev* 17, 400-407 (2007).
103. Chung, J.H., Bell, A.C. & Felsenfeld, G. Characterization of the chicken beta-globin insulator. *Proc Natl Acad Sci U S A* 94, 575-580 (1997).
104. Bell, A.C., West, A.G. & Felsenfeld, G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98, 387-396 (1999).
105. Ryu, B.Y., Persons, D.A., Evans-Galea, M.V., Gray, J.T. & Nienhuis, A.W. A chromatin insulator blocks interactions between globin regulatory elements and cellular promoters in erythroid cells. *Blood Cells MolDis* 39, 221-228 (2007).
106. Ryu, B.Y., Evans-Galea, M.V., Gray, J.T., Bodine, D.M., Persons, D.A. & Nienhuis, A.W. An experimental system for the evaluation of retroviral vector design to diminish the risk for proto-oncogene activation. *Blood* 111, 1866-1875 (2008).
107. Yao, S., Osborne, C.S., Bharadwaj, R.R., Pasceri, P., Sukonnik, T., Pannell, D., Recillas-Targa, F., West, A.G. & Ellis, J. Retrovirus silencer blocking by the cHS4 insulator is CTCF independent. *Nucleic Acids Res* 31, 5317-5323 (2003).
108. Nishino, T., Tubb, J. & Emery, D.W. Partial correction of murine beta-thalassemia with a gammaretrovirus vector for human gamma-globin. *Blood Cells MolDis* 37, 1-7 (2006).
109. Aker, M., Tubb, J., Groth, A.C., Bukovsky, A.A., Bell, A.C., Felsenfeld, G., Kiem, H.P., Stamatoyannopoulos, G. & Emery, D.W. Extended core sequences from the cHS4 insulator are necessary for protecting retroviral vectors from silencing position effects. *Hum Gene Ther* 18, 333-343 (2007).
110. Li, C.L. & Emery, D.W. The cHS4 chromatin insulator reduces gammaretroviral vector silencing by epigenetic modifications of integrated provirus. *Gene Ther* 15, 49-53 (2008).
111. Ma, Y., Ramezani, A., Lewis, R., Hawley, R.G. & Thomson, J.A. High-level sustained transgene expression in human embryonic stem cells using lentiviral vectors. *Stem Cells* 21, 111-117 (2003).

112. Chang, L.J., Liu, X. & He, J. Lentiviral siRNAs targeting multiple highly conserved RNA sequences of human immunodeficiency virus type 1. *Gene Ther* 12, 1133-1144 (2005).
113. Pluta, K., Luce, M.J., Bao, L., Agha-Mohammadi, S. & Reiser, J. Tight control of transgene expression by lentivirus vectors containing second-generation tetracycline-responsive promoters. *J Gene Med* 7, 803-817 (2005).
- 5 114. Jakobsson, J., Rosenqvist, N., Thompson, L., Barraud, P. & Lundberg, C. Dynamics of transgene expression in a neural stem cell line transduced with lentiviral vectors incorporating the cHS4 insulator. *Experimental cell research* 298, 611-623 (2004).
115. Leboulch, P., Huang, G.M., Humphries, R.K., Oh, Y.H., Eaves, C.J., Tuan, D.Y. & London, I.M. Mutagenesis of retroviral vectors transducing human beta-globin gene and beta-globin locus control region derivatives results in stable transmission of an active transcriptional structure. *EMBO J* 13, 3065-3076 (1994).
- 10 116. Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenko, V.V. & Ren, B. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 128, 1231-1245 (2007).
117. Yusufzai, T.M. & Felsenfeld, G. The 5'-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element. *Proc Natl Acad Sci U S A* 101, 8620-8624 (2004).
- 15 118. Phillips, J.E. & Corces, V.G. CTCF: master weaver of the genome. *Cell* 137, 1194-1211 (2009).
119. Giles, K.E., Gowher, H., Ghirlando, R., Jin, C. & Felsenfeld, G. Chromatin boundaries, insulators, and long-range interactions in the nucleus. *Cold Spring Harbor symposia on quantitative biology* 75, 79-85 (2010).
120. Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I. & Zhao, K. High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823-837 (2007).
- 20 121. Wang, H., Maurano, M.T., Qu, H., Varley, K.E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R., Thurman, R.E., Kaul, R., Myers, R.M. & Stamatoyannopoulos, J.A. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome research* 22, 1680-1688 (2012).
122. Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Goncalves, A., Kutter, C., Brown, G.D., Marshall, A., Flicek, P. & Odom, D.T. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* 148, 335-348 (2012).
- 25 123. Renda, M., Baglivo, I., Burgess-Beusse, B., Esposito, S., Fattorusso, R., Felsenfeld, G. & Pedone, P.V. Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. *J Biol Chem* 282, 33336-33345 (2007).
124. Dickson, J., Gowher, H., Strogantsev, R., Gaszner, M., Hair, A., Felsenfeld, G. & West, A.G. VEZF1 elements mediate protection from DNA methylation. *PLoS Genet* 6, e1000804 (2010).
- 30 125. Li, C.L., Xiong, D., Stamatoyannopoulos, G. & Emery, D.W. Genomic and functional assays demonstrate reduced gammaretroviral vector genotoxicity associated with use of the cHS4 chromatin insulator. *Mol Ther* 17, 716-724 (2009).
126. Lisowski, L. & Sadelain, M. Locus control region elements HS1 and HS4 enhance the therapeutic efficacy of globin gene transfer in beta-thalassemic mice. *Blood* 110, 4175-4178 (2007).
- 35 127. Nagel, R.L., Bookchin, R.M., Johnson, J., Labie, D., Wajcman, H., Isaac-Sodeye, W.A., Honig, G.R., Schiliro, G., Crookston, J.H. & Matsutomo, K. Structural bases of the inhibitory effects of hemoglobin F and hemoglobin A2 on the polymerization of hemoglobin S. *Proc Natl Acad Sci U S A* 76, 670-672 (1979).
128. Sadelain et al., *Proc. Nat'l Acad. Sci. (USA)* (1995);92:6728-6732.
- 40 129. Armstrong, J.A., Emerson, B.M., 1996. NFE2 disrupts chromatin structure at human fl-globin locus control region hypersensitive site 2 in vitro. *Mol. Cell. Biol.* 16, 5634-5644.
130. Caterina, J.J., Ciavatta, D.J., Donze, D., Behringer, R.R., Townes, T.M., 1994. Multiple elements in human fl-globin locus control region 5' HS2 are involved in enhancer activity and position-independent transgene expression. *Nucleic Acids Res.* 22, 1006 1011.
- 45 131. Moi, P., Kan, Y.W., 1990. Synergistic enhancement of globin gene expression by activator protein-1-like proteins. *Proc. Natl. Acad. Sci. USA* 87, 9000-9004.
132. Ney, P., Sorrentino, B., McDonagh, K., Nienhuis, A., 1990. Tandem AP-1-binding sites within the human β -globin dominant control region function as an inducible enhancer in erythroid cells. *Genes Dev.* 4, 993 1006.
- 50 133. Shivdasani, R.A., Rosenblatt, M.F., Zucker-Franklin, D., Jackson, C.W., Hunt, P., Saris, C.J.M., Orkin, S.H., 1995. Transcription factor NF-E2 is required for platelet formation independent of the actions of thrombopoietin/MGDF in megakaryocyte development. *Cell* 81,695-704.
134. Talbot, D., Grosveld, F., 1991. The 5'HS2 of the globin locus control region enhances transcription through the interaction of a multimeric complex binding at two functionally distinct NF-E2 binding sites. *EMBO J.* 10, 1391-1398.
- 55 135. Hardison et al., *Gene* (1997);205:73-94.
136. Elnitski et al., *The Journal of Biological Chemistry* (1997);272(1):369-378; Horak et al., *PNAS* (2002);99(5):2924-2929.
137. Shimotsuma et al., *Journal of Biological Chemistry* (2010);285(19):14495-14503.

[0223] From the foregoing description, it will be apparent that variations and modifications may be made to the presently disclosed subject matter described herein to adopt it to various usages and conditions. Such embodiments are also within the scope of the following claims.

[0224] All patents and publications and sequences referred to by accession or reference number mentioned in this specification are herein incorporated by reference to the same extent as if each independent patent and publication and sequence was specifically and individually indicated to be incorporated by reference.

5

10

15

20

25

30

35

40

45

50

55

EP 3 620 521 A1

SEQUENCE LISTING

<110> MEMORIAL SLOAN-KETTERING CANCER CENTER

<120> GLOBIN GENE THERAPY FOR TREATING HEMOGLOBINOPATHIES

<130> JA88015P.EPPD1

<140>

<141> 2015-09-04

<150> 15837740.8

<151> 2015-09-04

<10> PCT/US2015/048698

<151> 2015-09-04

<150> 62/045,997

<151> 2014-09-04

<160> 26

<170> PatentIn version 3.5

<210> 1

<211> 305

<212> DNA

<213> Homo sapiens

<400> 1

tccttccttt ctaaagacg agagagacag aagaattcct caaggtagt gtgtccagca 60

tgcaaccttt ccttcctgga tgagcatccc tggagtagga gagccagcct gcctcctgcg 120

ctggcacaga gcccggttcc ctagacaact gcctctccaa atctgatgtc cagcgccacc 180

tggtgtccac atcaagcaga cacaattaat agtcaacctg ttcaggaaaa ctgtgagggg 240

gaaaaaaaaag aaagaggatt tatgaaggga aaagaaagtt tagaggatat gccacgattg 300

gctag 305

<210> 2

<211> 1074

<212> DNA

<213> Homo sapiens

<400> 2

aagtaaaactt ccacaaccgc aagcttattg aggctaaggc atctgtgaag gaaagaaaca 60

tctcctctaa accactatgc tgctagagcc tcttttctgt actcaagcct cattcagaca 120

ctagtgtcac cagtctcctc atatacctat tgtattttct tcttcttgct ggtttagtca 180

tgttttctgg gagcttaggg gcttatttta ttttgttttg ttttctaatac aacagagatg 240

ggcaaaccga ttattttttt ctttagactt gggatggtga tagctgggca gcgtcagaaa 300

ctgtgtgtgg atatagataa gagctcggac tatgctgagc tgtgatgagg gagggaccta 360

gccaaaggca gtgagagtca gaatgctcct gctattgcct tctcagtcct cagccttggt 420

EP 3 620 521 A1

	ttctacacaa gtagatacat agaaaaggct atagggttagt gtttgagagt cctgcatgag	480
	ttagttgctc agaaatgccc gataaatatg ttatgtgtgt ttatgtatat atatgtttta	540
5	tatatatata tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg ttacaaatat gtgattatca	600
	tcaaaacgtg agggctaaag tgaccagata acttgcaggt cctaggatac caggaaaata	660
	aattacattc caaaaattta actgagactt taaaaaaaaa aaaaaaaaaa aaaaaaaac	720
10	cagtgatcca tggacacagg gaggggaaca tcacacactg gggcctgttg ggggtggggg	780
	gctaggggaa ggatagcatt aggagaaata cctaagttag atgacgggtt gatgggtgca	840
	gcaaacacc accatggcacatg taccacagaa cttaaagcat attaaaaaa cagtgatcat	900
15	aaaagaagct caaatttaac tataagagac ggaatggctc ccacaattct taactataat	960
	cttacagaat attctcattg aatagaagta tgcttatcat tagagatttg gacagccagg	1020
20	aaagcacaga aaaaaaaaaa aggagctctg ttgccttata gcctagaggt gttt	1074
	<210> 3	
	<211> 602	
	<212> DNA	
25	<213> Homo sapiens	
	<400> 3	
	ggcatctgtg aaggaaagaa acatctcctc taaaccacta tgctgctaga gcctcttttc	60
	tgtactcaag cctcattcag acactagtgt caccagtctc ctcatatacc tattgtattt	120
30	tcttcttctt gctggtttag tcatgttttc tgggagctta ggggcttatt ttattttggt	180
	ttgttttcta atcaacagag atgggcaaac ccattatttt tttctttaga cttgggatgg	240
	tgatagctgg gcagcgtcag aaactgtgtg tggatataga taagagctcg gactatgctg	300
35	agctgtgatg agggagggac ctagccaaag gcagtgagag tcagaatgct cctgctattg	360
	ccttctcagt cccacgctt ggtttctaca caagtagata catagaaaag gctatagggt	420
40	agtgtttgag agtcctgcat gagttagtgt ctcagaaatg cccgataaat atgttatgtg	480
	tgtttatgta tatatatgtt ttatatatat atatgtgtgt gtgtgtgtgt gtgtgtgtg	540
	tgtttacaaa tatgtgatta tcatcaaac gtgagggcta aagtgaccag ataacttgca	600
45	gg	602
	<210> 4	
	<211> 489	
	<212> DNA	
50	<213> Homo sapiens	
	<400> 4	
	ggcatctgtg aaggaaagaa acatctcctc taaaccacta tgctgctaga gcctcttttc	60
55	tgtactcaag cctcattcag acactagtgt caccagtctc ctcatatacc tattgtattt	120
	tcttcttctt gctggtttag tcatgttttc tgggagctta ggggcttatt ttattttggt	180

EP 3 620 521 A1

	ttgttttcta atcaacagag atgggcaaac ccattatttt tttctttaga cttgggatgg	240
	tgatagctgg gcagcgtcag aaactgtgtg tggatataga taagagctcg gactatgctg	300
5	agctgtgatg agggagggac ctagccaaag gcagtgagag tcagaatgct cctgctattg	360
	ccttctcagt ccccacgctt ggtttctaca caagtagata catagaaaag gctatagggt	420
	agtgtttgag agtcctgcat gagttagttg ctcagaaatg cccgataaat atgttatgtg	480
10	tgtttatgt	489
	<210> 5	
	<211> 1301	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 5	
	aagctttcat taaaaaaaaagt ctaaccagct gcattcgact ttgactgcag cagctggtta	60
20	gaaggttcta ctggaggagg gtcccagccc attgctaaat taacatcagg ctctgagact	120
	ggcagtatat ctctaacagt ggttgatgct atcttctgga acttgccctgc tacattgaga	180
	ccactgaccc atacatagga agcccatagc tctgtcctga actgttaggc cactgggtcca	240
25	gagagtgtgc atctcctttg atcctcataa taaccctatg agatagacac aattattact	300
	cttactttat agatgatgat cctgaaaaca taggagtcaa ggcacttgcc cctagctggg	360
30	ggtatagggg agcagtccca tgtagtagta gaatgaaaaa tgctgctatg ctgtgcctcc	420
	cccacctttc ccatgtctgc cctctactca tgggtctatct ctctgggctc ctgggagtca	480
	tggactccac ccagcaccac caacctgacc taaccaccta tctgagcctg ccagcctata	540
35	acccatctgg gccctgatag ctggtggcca gccctgaccc caccaccacc tccctggaac	600
	ctctgataga cacatctggc acaccagctc gcaaagtcac cgtgaggggc ttgtgtttgc	660
	tgagtcaaaa ttccttgaaa tccaagtcct tagagactcc tgctcccaa tttacagtca	720
40	tagacttctt catggctgtc tcctttatcc acagaatgat tcctttgctt cattgccccca	780
	tccatctgat cctcctcatc agtgcagcac agggcccatg agcagttagct gcagagtctc	840
	acataggtct ggcactgcct ctgacatgtc cgaccttagg caaatgcttg actcttctga	900
45	gctcagtctt gtcatggcaa aataaagata ataatagtgt ttttttatgg agttagcgtg	960
	aggatggaaa acaatagcaa aattgattag actataaaag gtctcaacaa atagtagtag	1020
50	attttatcat ccattaatcc ttccctctcc tctcttactc atcccatcac gtatgcctct	1080
	taattttccc ttacctataa taagagttat tcctcttatt atattcttct tatagtgatt	1140
	ctggatatta aagtgggaat gaggggcagg ccactaacga agaagatgtt tctcaaagaa	1200
55	gccattctcc ccacatagat catctcagca ggggttcagga agataaagga ggatcaaggt	1260
	cgaaggtagg aactaaggaa gaacactggg caagtggatc c	1301

EP 3 620 521 A1

<210> 6
<211> 1065
<212> DNA
<213> Homo sapiens

5
5
10
15
20
25
30
35
40

<400> 6
tgagcccctt ttcctctaac tgaaagaagg aaaaaaaaaa tggaacccaa aatattctac 60
atagtttcca tgtcacagcc agggctgggc agtctcctgt tatttctttt aaaataaata 120
tatcatttaa atgcataaat aagcaaaccg tgctcgaggaa tgggagggag agtctctgga 180
gtccaccctt tctcgccctt ggctctgcag atagtgtctat caaagccctg acagagccct 240
gcccattgct gggccttgga gtgagtcagc ctagtagaga ggcagggcaa gccatctcat 300
agctgctgag tgggagagag aaaagggctc attgtctata aactcaggctc atggctattc 360
ttattctcac actaagaaaa agaattgagat gtctacatat accctgcgtc ccctcttggtg 420
tactggggcc cccaagagct ctctaaaagt gatggcaaag tcattgctgct agatgccatc 480
ccatctatta taaacctgca tttgtctcca cacaccagtc atggacaata accctcctcc 540
caggtccacg tgcttgtctt tgtataatac tcaagtaatt tcggaaaatg tattctttca 600
atcttggttct gttattcctg tttcaatggc ttagtagaaa aagtacatac ttgttttccc 660
ataaattgac aatagacaat ttcacatcaa tgtctatatg ggctcgttggtg tttgctgtgt 720
ttgcaaaaac tcacaataac tttatatattgt tactactcta agaaagttac aacatgggtga 780
atacaagaga aagctattac aagtccagaa aataaaagtt atcatcttga ggcctcagct 840
ttctaggaat aatatcaata ttacaaaatt taatctaaca attatgaaca gcaatgagat 900
aatatgtaca aagtaccag acctatgtgg tagagcatca aggaagcgca ttgcggagca 960
gttttttgtt tgtttgtttt tgtattctgt ttogtgaggc aaggtttcac tctgctgtcc 1020
aggctggagt gcagtggcaa gatcatgtct cactgcagcc ttgac 1065

<210> 7
<211> 1065
<212> DNA
<213> Homo sapiens

45
50
55

<400> 7
tgagcccctt ttcctctaac tgaaagaagg aaaaaaaaaa tggaacccaa aatattctac 60
atagtttcca tgtcacagcc agggctgggc agtctcctgt tatttctttt aaaataaata 120
tatcattaaa tgcataaata agcaaaccct gctcggggaat gggagggaga gtctctggag 180
tccaccctt ctcgccctg gctctgcaga tagtgctatc aaagccctga cagagccctg 240
cccattgctg ggccttgag tgagtcagcc tagtagagag gcagggcaag ccatctcata 300
gctgctgagt gggagagaga aaagggctca ttgtctataa actcagggtca tggctattct 360

EP 3 620 521 A1

	tattctcaca ctaagaaaaa gaatgagatg tctacatata ccctgcgtcc cctcttgtgt	420
	actggggccc ccaagagctc tctaaaagtg atggcaaagt cattgcgcta gatgccatcc	480
5	catctattat aaacctgcat ttgtctccac acaccagtca tggacaataa ccctcctccc	540
	aggtccacgt gcttgtcttt gtataatact caagtaattt cggaaaatgt attctttcaa	600
	tcttgttctg ttattcctgt ttcaatggct tagtagaaaa agtacatact tgttttccca	660
10	taaattgaca atagacaatt tcacatcaat gtctatatgg gtcgttgtgt ttgctgtgtt	720
	tgcaaaaact cacaataact ttatattgtt actactctaa gaaagttaca acatggtgaa	780
	tacaagagaa agctattaca agtccagaaa ataaaagtta tcatcttgag gcctcagctt	840
15	tctaggaata atatcaatat tacaaaatta atctaacaat tatgaacagc aatgagataa	900
	tatgtacaaa gtaccagac ctatgtggtg gagcatcaag gaagcgcatt gcggagcagt	960
20	tttttgtttg tttgtttttg tattctgttt cgtgaggcaa ggtttcactc tgctgtccag	1020
	gctggagtgc agtggcaaga tcatgtctca ctgcagcctt gacac	1065
	<210> 8	
	<211> 446	
25	<212> DNA	
	<213> Homo sapiens	
	<400> 8	
30	tggaacccaa aatattctac atagtttcca tgtcacagcc agggctgggc agtctcctgt	60
	tatttctttt aaaataaata tatcatttaa atgcataaat aagcaaacc cgtctcgggaa	120
	tgggagggag agtctctgga gtccaccctt tctcggccct ggctctgcag atagtgtctat	180
35	caaagccctg acagagccct gccattgtct gggccttgga gtgagtcagc ctagtagaga	240
	ggcagggcaa gccatctcat agctgctgag tgggagagag aaaagggtc attgtctata	300
	aactcaggtc atggctattc ttattctcac actaagaaaa agaattgagat gtctacatat	360
40	accctgcgtc ccctcttctg tactggggtc cccaagagct ctctaaaagt gatggcaaag	420
	tcattgcgtc agatgccatc ccatct	446
	<210> 9	
45	<211> 860	
	<212> DNA	
	<213> Homo sapiens	
	<400> 9	
50	gtatatgtgt atatatatat atatatattc aggaaataat atattctaga atatgtcaca	60
	ttctgtctca ggcatccatt ttctttatga tgccgtttga ggtggagttt tagtcagggtg	120
	gtcagcttct cctttttttt gccatctgcc ctgtaagcat cctgctgggg acccagatag	180
55	gagtcatcac tctaggctga gaacatctgg gcacacaccc taagcctcag catgactcat	240
	catgactcag cattgctgtg cttgagccag aagggttgct tagaaggtta cacagaacca	300

EP 3 620 521 A1

	gaaggcgggg gtggggcact gaccccgaca ggggcctggc cagaactgct catgcttgga	360
	ctatgggagg tcactaatgg agacacacag aaatgtaaca ggaactaagg aaaaactgaa	420
5	gcttatttaa tcagagatga gatgctggaa gggatagagg gagctgagct tgtaaaaagt	480
	atagtaatca ttcagcaaat ggttttgaag cacctgctgg atgctaaaca ctattttcag	540
	tgcttgaatc ataaataaga ataaaacatg tatcttattc ccacaagag tccaagtaaa	600
10	aaataacagt taattataat gtgctctgtc cccaggctg gagtgcagtg gcacgatctc	660
	agctcactgc aacctccgcc tcccgggttc aagcaattct cctgcctcag ccacccta	720
	agctgggatt acaggtgcac accaccatgc caggctaatt tttgtacttt ttgtagaggc	780
15	agggatatcac catgttgtcc aagatggtct tgaactcctg agtccaagc agtccacca	840
	cctcagcctc ccaaagtgtc	860
20	<210> 10 <211> 613 <212> DNA <213> Homo sapiens	
25	<400> 10 aagcaataga tggctctgcc ctgactttta tgcccagccc tggctcctgc cctccctgct	60
	cctgggagta gattggccaa ccctaggggtg tggctccaca gggtgaggtc taagtgatga	120
30	cagccgtacc tgtccttggc tcttctggca ctggcttagg agttggactt caaacctca	180
	gccctccctc taagatatat ctcttggccc cataccatca gtacaaattg ctactaaaa	240
	catcctcctt tgcaagtgt tttacgtaat atttggaatc acagcttggg aagcatattg	300
35	aagatcggtt tcccaatttt ctattacac aaataagaaa ttgatgcact aaaagtggaa	360
	gagttttgtc taccataatt cagctttggg atatgtagat ggatctcttc ctgcgtctcc	420
	agaatatgca aaatacttac aggacagaat ggatgaaaac tctacctcag ttctaagcat	480
40	atcttctcct tatttggatt aaaaccttct ggtaagaaaa gaaaaaaaaat atatatatat	540
	atgtgtatat atacacacat acatatacat atatatgcat tcatttgttg ttgtttttct	600
45	taatttgctc atg	613
50	<210> 11 <211> 265 <212> DNA <213> Homo sapiens	
	<400> 11 aagcaataga tggctctgcc ctgactttta tgcccagccc tggctcctgc cctccctgct	60
55	cctgggagta gattggccaa ccctaggggtg tggctccaca gggtgaggtc taagtgatga	120
	cagccgtacc tgtccttggc tcttctggca ctggcttagg agttggactt caaacctca	180

EP 3 620 521 A1

gccctccctc taagatatat ctcttgccc cataccatca gtacaaattg ctactaaaaa 240
catcctcctt tgcaagtgtt ttac 265

5
<210> 12
<211> 879
<212> DNA
<213> Homo sapiens

10
<400> 12
taggtattga ataagaaaaa tgaagttaag gtggttgatg gtaacactat gctaataact 60
gcagagccag aagcaccata agggacatga taaggaggcc agcagacctc tgatctcttc 120
15 ctgaatgcta atcttaaaca tcctgaggaa gaatgggact tccatttggg gtgggcctat 180
gatagggtaa taagacagta gtgaatatca agctacaaaa agcccccttt caaattcttc 240
tcagtcctaa cttttcatac taagcccagt ccttccaaag cagactgtga aagagtgtga 300
20 gttccgggag actagcactg cagattccgg gtcactgtga gtgggggagg cagggaagaa 360
gggctcacag gacagtcaaa ccatgcccc tgtttttcct tcttcaagta gacctctata 420
agacaacaga gacaactaag gctgagtggc caggcgagga gaaaccatct cgccgtaaaa 480
25 catggaagga acacttcagg ggaaaggtgg tatctctaag caagagaact gagtggagtc 540
aaggctgaga gatgcaggat aagcaaatgg gtagtgaaaa gacattcatg aggacagcta 600
aaacaataag taatgtaaaa tacagcatag caaaacttta acctccaaat caagcctcta 660
30 cttgaatcct tttctgaggg atgaataagg cataggcatc aggggctgtt gccaatgtgc 720
attagctgtt tgcagcctca cttctttca tggagtttaa gatatagtgt attttcccaa 780
ggtttgaact agctcttcac ttctttatgt tttaaatgca ctgacctccc acattccctt 840
35 tttagtaaaa tattcagaaa taatttaaata acatcattg 879

40
<210> 13
<211> 152
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
45 polynucleotide

<400> 13
tctcccacgc cctggtctca gcttggggag tggtcagacc ccaatggcga taaactctgg 60
caactttatc tgtgactgc aggtcagcc ccaacagctt tagctttcac aagcaggcag 120
50 gggaagggaa acacatatct ccagatatga gg 152

55
<210> 14
<211> 157
<212> DNA
<213> Artificial Sequence

EP 3 620 521 A1

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

5 <400> 14
ctaaaccct cccccaccct agccccaagc ttcatcttag ctccactcct gaccctatcc 60
agctaaaggt cccccaccag ctcttgcta tctagtcatt gcatatggca agacttgaaa 120
10 gtcctatctc aaagcagcag aattatcagc tacgact 157

<210> 15

<211> 141

<212> DNA

15 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

20 <400> 15
ccatccccca gcactccctg cccccacagc ccagacttga ccaactccca gctccgcctg 60
ggacttccag atatggggcc ccacccttgc aggccttggg gacgctgaag atattgacta 120
25 tctgcgtgcc ggaaaagggt g 141

<210> 16

<211> 171

<212> DNA

30 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

35 <400> 16
aaaggctggg ggtgggagta gcggatttga agcaactgtt ggcctacaga ggtgtggcaa 60
gcagagcacc tcagaactca ggcgtactgc ccgccgccg agccctgcga gggccgatat 120
40 cgagggtgtg gcccttatct gcaccagca gagcgccggc ggggtacggt c 171

<210> 17

<211> 195

<212> DNA

45 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

50 <400> 17
cagttgcctc agctgagtat gtcttctaaa gataatgtcg attgtgtatg gctgatggga 60
ttctaggacc aagcaagagg tttttttttt tccccacat acttaacggt tctatatttc 120
55 tatttgaatt cgactggaca gttccatttg aattatttct ctctctctct ctctctgaca 180

catttttatct tgcca

195

<210> 18

<211> 14

<212> DNA

<213> Homo sapiens

<400> 18

caccagggtgg cgct

14

<210> 19

<211> 81706

<212> DNA

<213> Homo sapiens

<400> 19

ggatcctcac atgagttcag tatataattg taacagaata aaaaatcaat tatgtattca 60

agttgctagt gtcttaagag gttcacattt ttatctaact gattatcaca aaaatacttc 120

gagttacttt tcattataat tcctgactac acatgaagag actgacacgt aggtgcctta 180

cttaggtagg ttaagtaatt tatccaaaac cacacaatgt agaacctaaag ctgattcggc 240

catagaaaca caatatgtgg tataaatgag acagagggat ttctctcctt cctatgctgt 300

cagatgaata ctgagataga atatttagtt catctatcac acattaaacg ggactttaca 360

tttctgtctg ttgaagattt ggggtgtggg ataactcaag gtatcatatc caagggatgg 420

atgaaggcag gtgactctaa cagaaagga aaggatgttg gcaaggctat gttcatgaaa 480

gtatatgtaa aatccacatt aagcttcttt ctgcatgcat tggcaatgtt tatgaataat 540

gtgtatgtaa aagtgtgctg tatattcaaa agtggttcat gtgcctaggg gtgtcaaata 600

ctttgagttt gtaagtatat acttctctgt aatgtgtctg aatatctcta tttacttgat 660

tctcaataag taggtatcat agtgaacatc tgacaaatgt ttgaggaaca atttagtggt 720

tacctattca ccaaaattta ttaaatgcct aatctgtatc agatatacaa ttatctggcg 780

aaatctgtaa ttcctaattt aaacagctgt gtagcctaata tagggataaa ggcattgcaaa 840

cccataattt gtgtaggttg aaatgagcta tagaaaaatg cagtatatat atcagaagtc 900

tttaggggtca tgaaaaggaa tgggtcaactg aactgccag ggactcatat gtaagagata 960

actaatgtga agtgacttta aaggagaaat tagcagaagt tttctttcca tgtctcctca 1020

tcatgtttaca ataacggaag agattaaaac aacaaatata tttagacagc aatgtttatc 1080

ctgggttagat gtttttaatct aaatctatct tggagtgtta aaatgcattt gctcacctac 1140

tttaaaatat aaatgaaggt aggaacctgt agatacaaaa agttggagaa aaaaagacaa 1200

taaagatgac aaaaatctat taatccttga tagaaaatga gaagagataa aacactgggt 1260

tacataaaga aaataagatg gatagatagc agatccttat aaaagtgata atttgagaaa 1320

aaaaatactc catattctga gtttcttcac ataaaataat acaaatctgc tgtggttaagt 1380

EP 3 620 521 A1

	tacaaagaga tagatTTTTT atcattatat aaaagatatt ttaaacagag ttatacaaca	1440
	aaggaacaga ctatgtcata tattctcact tatcactata aacatctcag aaaaatctgc	1500
5	aaaatcattt catagcattt taaatagtta ggaataatgt agaaaactga aacagttcta	1560
	agtttcccac aaacttagag tctcaaagt tgcattacct aacttacctg caaatatttt	1620
	atacaaattt gcacatgcta ctctagtcaa aaatatatgt acattatggg tattttctgt	1680
10	gtgtaacttg gttctagttg cttctttcag aaatagcctc tatttttgat ttacctgata	1740
	aaatcacatt cctctccaaa gccttctaaa tacttccaga ctaactactt tttagtacat	1800
	ctaagaagaa aagagttttg tctcttatcc acctctgagt caaaaagcag catgtccatc	1860
15	aattggtaca tagttcccac agccccactt agctctggat tggagttcta cttggcattg	1920
	tttgcaacta catggacgta aaatgcatgg attctcttga aaaaatgttt ctgccatgat	1980
20	gttctctgaa agagactaac cttccctcgc tttgcagaga aagactcgtg taatccttga	2040
	caatgtcatc tcatctattt attcccatgt ctacccatat gtgaccttca tgtctttgct	2100
	ctaagcccct acatcctcaa tctacacact aggatagtat aaaagtaata gtaataatag	2160
25	tagtaatagt aataacaata caatgattat ggcttatact atacacaaga cactgttgat	2220
	atattatttc atttagtatt cacagtaact ctgtgcctca agtactattg taataccctt	2280
	taagaggagg aaactgaggc acagggccct aaagtaatat tccaagatga agtggctact	2340
30	aactgacaga gggcataatt caactcatga tatttggtct tagaatacat gctctgaatc	2400
	attatacaat aataattcat gaggaacat tttttaaagc ctaagttatt tgctctgaaa	2460
35	taagacataa tttggggtga gaaagcttag attccatgaa gtattacagc atttggtagt	2520
	ctttttgcac tccaggtctt atttttactg cttaaacata ataaaacata tggttcagta	2580
	tgcttttgat tttaacaata tattcctgtt atttttggaa gcacaggggtg tgggataatg	2640
40	ctaattacta gtgattagta ttgagaggtg acagcgtgct ggcagtcctc acagccctcg	2700
	ctcgtctctg gcgcctcctc tgcttgggt cccacattgg tggcacttga ggagcccttc	2760
	agccggccgc tgcactgtgg gagccctttt ctgggctggc caaggccaga gccggctccc	2820
45	tcagcttgcc aggaggtgtg gagggacaga cgcgggcagg aaccgggctg tgcgcctgc	2880
	ttgagggagt tccgggtggg catgggctcc gaggaccccg cactcggagc cgccagccgg	2940
	ccccaccggc cgcgggcagt gaggggctta gcacctgggc cagcagctgc tgtgctcaat	3000
50	tcctcgccgg gccttagctg ccttcctgcg gggcagggct cgggacctgc agcgcgccat	3060
	gcctgagcct cccacacctc atgggctcct gtgcggcccg agcctcgccg acgagcgccg	3120
55	ccccctgctc cagggcaccc agtcccatcg accaccaag ggctgaagag tgcgggcgca	3180
	cggcagggga ctggcaggca gctccccctg cagcccaggt gcgggatcca ctgggtgaag	3240

EP 3 620 521 A1

	ccggctaggc	tcctgagttt	gctggggatg	cgaagaaccc	ttatgtctag	ataagggatt	3300
	gtaaatacac	caattggcac	tctgtatcta	gctcaagggt	tgtaaacaca	ccaatcagca	3360
5	ccctgtgtct	agctcagggt	ttgtgaatgc	accaatcaac	actctatcta	gctactctgg	3420
	tggggccttg	gagaaccttt	atgtctagct	cagggattgt	aaatacacca	atcggcagtc	3480
	tgtatctagc	tcaaggtttg	taaacacacc	aatcagcacc	ctgtgtctag	ctcagggttt	3540
10	gtgaatgcac	caatcaacac	tctgtatcta	gctactctgg	tggggacgtg	gagaaccttt	3600
	atgtctagct	cagggattgt	aaatacacca	ctcggcagtc	tgtatctagc	tcaaggtttg	3660
	taaacacacc	aatcagcacc	ctgtgtctag	ctcagggttt	gtgaatgcac	caatcaacac	3720
15	tctgtatcta	gctactctgg	tggggacttg	gagaaccttt	gtgtggacac	tctgtatcta	3780
	gctaactctg	tggggacgtg	gagaaccttt	gtgtctagct	catggattgt	aaatgcacca	3840
	atcagtgcc	tgtcaaaaca	gaccactggg	ctctaccaat	cagcaggatg	tgggtggggc	3900
20	cagataagag	aataaaagca	ggctgcccg	gccagcagtg	gcaaccgct	cgggtcccct	3960
	tccacactgt	ggaagctttg	ttctttcgct	ctttgcaata	aatcttgctg	ctgctcactg	4020
	tttgggtcta	cactgccttt	atgagctgta	acgtccaccg	cgaaggtctg	cagcttcact	4080
25	cttgaagcca	gcgagaccac	gaaccaccg	ggaggaacga	acaactccag	aggcgcggcc	4140
	ttaagagctg	gaacgttcac	tgtgaaggtc	tgcagcttca	ctcctgagcc	agcgagacca	4200
	cgaacccatc	agaaggaaga	aactccgaac	acatccaaac	atcagaacga	acaaactcca	4260
30	cacacgcagc	ctttaagaac	tgtaacactc	accacgaggg	tccccggctt	cattcttgaa	4320
	gtcagtgaaa	ccaagaaccc	accaattccg	gacacagtat	gtcagaaaca	atatgagtca	4380
35	ctaaatcaat	atactttctca	acaatttcca	acagccottg	caattaactt	ggccatgtga	4440
	ctggttgtga	ctaaaataat	gtggagataa	taatgtgtta	ctccctaagg	cagagtgtccc	4500
	ttctatcatt	ctctttccct	tcctctatgt	ggcagaaagt	aaaagattct	gaaatgataa	4560
40	agtcaatcac	aggaaggcac	ctggactcct	ggcccactgc	ttggaggaga	gcactcagga	4620
	ccatgaacat	ctgactgtga	cgtagcaata	aagaaaccca	cgtttcatat	gaaactgctt	4680
	aaaattaatg	gcacaagtca	tgtttttgat	gttgcacatt	tgtctttatt	tgtggcttgt	4740
45	tttgcttcca	catcaatcca	ctcaaggcct	acattctgct	ataatgcaat	ttcaagttct	4800
	ttacaggccg	agaaaaatga	atctgaattc	ctgacctcca	aaagtgatca	agatattttt	4860
	agttcaggct	ccaaaatttt	ctcattttca	taggttttcc	tcgattgatc	attattcatg	4920
50	atttgcaagg	aatcattcaa	tgttttctaa	atctattact	gcacccctgac	acatatgaca	4980
	ttttaactat	gttccagatt	tttgaatgaa	gagtgtaaat	tttaaatgtt	ttcaccacaa	5040
55	aaaataagta	tgtgaagtgg	tggatttgtt	aattagcctt	atttaaccat	ttaatattgt	5100
	acacgtacac	caaagcatca	tgttgtaccc	catgaataca	cacaattatt	atttgtcaat	5160

EP 3 620 521 A1

	ttaaaatgaa ataataaaaa ataacaaagg cattagcctc tgcattgcct ttaccgggtca	5220
	tcctcacggt gactaacgca aaaaacgttc tatttcatcc ttacaaacat ccctatcttt	5280
5	gatgcctctt tgtctagatc tctatcccct cctgttttct ctacgttatt tatatgggta	5340
	tcatcaccat cctggacaac atcaggacag atatccctca ccaagccaat gttcctctct	5400
	atgttggtc aaatgtcctt gaactttcct ttcaccaccc ttccacagc caaaaggata	5460
10	ttgtagttta atgcctcaga gttcagcttt taagcttctg acaaattatt ctctctcttt	5520
	aggttctcct ttatggaatc ttctgtactg atggccatgt cctttaacta ctatgtagat	5580
	atctgtact acctgtatta tgccctctacc ttatttagca gagttatctg tactgttggc	5640
15	atgacaatca ttgtttaata tgacttgcct ttcccttttc tgctattctt gatcaaatgg	5700
	ctcctctttc ttgtctctct catttctcct gccttcaactt ggacgtgctt cacgtagtct	5760
	gtgcttatga ctggattaaa aattgatatg gacttatcct aatgttggtc gtcataatat	5820
20	gggttttatg gtccattatt atttccctatg cattgatctg gagaaggctt caatcctttt	5880
	actctttgtg gaaaatatct gtaaacccttc tggttcactc tgctatagca atttcagttt	5940
25	aggctagtaa gcatgaggat gcctccttct ctgatttttc ccacagtctg ttggtcacag	6000
	aataacctga gtgattactg atgaaagagt gagaatgtta ttgatagtca caatgacaaa	6060
	aaacaaacaa ctacagtcaa aatgtttctc tttttattag tggattatat ttcttgacct	6120
30	atatctggca ggactcttta gagaggtagc tgaagctgct gttatgacca ctagagggaa	6180
	gaagatacct gtggagctaa tgggtccaaga tgggtggagcc ccaagcaagg aagttgttaa	6240
	ggagcccttt tgattgaagg tgggtgcccc caccttacag ggacaggaca tctggatact	6300
35	cctcccagtt tctccagttt ccctttttcc taatatatct cctgataaaa tgtctatact	6360
	cacttcccca tttctaataa taaagcaaag gctagttagt aagacatcac cttgcatttt	6420
40	gaaaatgcc tagactttca aaattatttc atacatcggc ctttctttat ttcaagagtc	6480
	cagaaatggc aacattacct ttgattcaat gtaatggaaa gagctctttc aagagacaga	6540
	gaaaagaata atttaatttc tttccccaca cctccttccc tgtctcttac cctatcttcc	6600
45	ttccttctac cctccccatt tctctctctc atttctcaga agtatatttt gaaaggattc	6660
	atagcagaca gctaaggctg gttttttcta agtgaagaag tgatattgag aaggtagggt	6720
	tgcatgagcc ctttcagttt tttagtttat atacatctgt attgttagaa tgttttataa	6780
50	tataaataaa attatttctc agttatatac tagctatgta acctgtggat atttccttaa	6840
	gtattacaag ctatacttaa ctcaacttga aaactcaa ataacctgc ttcatagtta	6900
	ttaataagga ttaagtgaga taatgccc atagattccta ttaataacag ataaatacat	6960
55	acacacacac acacattgaa aggattctta ctttgtgcta ggaactataa taagttcatt	7020

EP 3 620 521 A1

	gatgcattat atcattaagt tctaatttca acactagaag gcaggtatta tctaaatttc	7080
	atactggata cctccaaact cataaagata attaaattgc cttttgtcat atattttattc	7140
5	aaaagggtaa actcaaacta tggccttgtct aattttatat atcaccctac tgaacatgac	7200
	cctattgtga tattttataa aattattctc aagttattat gaggatgttg aaagacagag	7260
	aggatgggggt gctatgcccc aaatcagcct cacaattaag ctaagcagct aagagtcttg	7320
10	cagggtagtg tagggaccac agggttaagg gggcagtaga attatactcc cacttttagtt	7380
	tcatttcaaa caatccatac acacacagcc ctgagcactt acaaattata ctacgctcta	7440
	tactttttgt ttaaagtgtat aaataagtgg atgaaagaat agatagatag atagacagat	7500
15	agatgataga tagaataaat gcttgccttc atagctgtct ccctaccttg ttcaaaatgt	7560
	tcctgtccag accaaagtac cttgccttca cttaagtaat caattcctag gttatattct	7620
	gatgtcaaag gaagtcaaaa gatgtgaaaa acaatttctg acccacaact catgctttgt	7680
20	agatgactag atcaaaaaat ttcagccata tottaacagt gagtgaacag gaaatctcct	7740
	cttttcccta catctgagat ccagcttctc aagaccttca attctcactc ttgatgcaac	7800
25	agaccttgga agcatacagg agagctgaac ttggtcaaca aaggagaaaa gtttgttggc	7860
	ctccaaaggc acagctcaaa cttttcaagc cttctctaata cttaaaggta aacaagggtc	7920
	tcattttctt gagaacttca gggaaaatag acaaggactt gcctgggtgct tttggtaggg	7980
30	gagcttgcac tttccccctt tctggaggaa atattttatcc ccaggtagtt ccctttttgc	8040
	accagtgggt ctttgaagag acttccacct gggaacagtt aaacagcaac tacagggcct	8100
	tgaactgcac actttcagtc cggtcctcac agttgaaaag acctaagctt gtgcctgatt	8160
35	taagcctttt tgggtcataaa acattgaatt ctaatctccc tctcaaccct acagtcaccc	8220
	at ttggtata ttaaagatgt gttgtctact gtctagtatc cctcaagtag tgtcaggaat	8280
	tagtcattta aatagtctgc aagccaggag tgggtggctca tgtctgtaat tccagcactt	8340
40	gagaggtaga agtgggagga ctgcttgagc tcaagagttt gatattatcc tggacaacat	8400
	agcaagacct cgtctctact taaaaaaaaa aaaaaaatta gccaggcatg tgatgtacac	8460
45	ctgtagtccc agctactcag gaggccgaaa tgggaggatc ccttgagctc aggaggtcaa	8520
	ggctgcagtg agacatgac ttgccactgc actccagcct ggacagcaga gtgaaacctt	8580
	gcctcacgaa acagaataca aaaacaaaaca aacaaaaaac tgctccgcaa tgcgcttctt	8640
	tgatgtctta ccacataggt ctgggtactt tgtacacatt atctcattgc tgttcataat	8700
50	tgttagatta attttgtaat attgatatta ttcctagaaa gctgaggcct caagatgata	8760
	actttttatt tctggacttg taatagcttt ctcttgtatt caccatgttg taactttctt	8820
55	agagtagtaa caatataaag ttattgtgag tttttgcaaa cacagcaaac acaacgaccc	8880
	atatagacat tgatgtgaaa ttgtctattg tcaatttatg ggaaaacaag tatgtacttt	8940

EP 3 620 521 A1

	ttctactaag ccattgaaac aggaataaca gaacaagatt gaaagaatac attttccgaa	9000
	attacttgag tattatacaa agacaagcac gtggacctgg gaggaggggtt attgtccatg	9060
5	actggtgtgt ggagacaaat gcaggtttat aatagatggg atggcatcta gcgcaatgac	9120
	tttgccatca ctttttagaga gctcttgggg accccagtac acaagagggg acgcagggtta	9180
	tatgtagaca tctcattctt tttcttagtg tgagaataag aatagccatg acctgagttt	9240
10	atagacaatg agcccttttc tctctcccac tcagcagcta tgagatggct tgccctgcct	9300
	ctctactagg ctgactcact ccaaggccca gcaatgggca gggctctgtc agggctttga	9360
	tagcactatc tgcagagcca gggccgagaa ggggtggact ccagagactc tccctcccat	9420
15	tcccgcagcag gggttgctta tttatgcatt taaatgatat atttatttta aaagaaataa	9480
	caggagactg ccagaccctg gctgtgacat ggaaactatg tagaatattt tgggttccat	9540
20	ttttttttcc ttctttcagt tagaggaaaa ggggtcact gcacatacac tagacagaaa	9600
	gtcaggagct ttgaatccaa gcctgatcat ttccatgtca tactgagaaa gtccccaccc	9660
	ttctctgagc ctcagtttct ctttttataa gtaggagtct ggagtaaagt atttccaatg	9720
25	gctctcattt caatacaaaa tttccgttta ttaaatgcat gagcttctgt tactccaaga	9780
	ctgagaagga aattgaacct gagactcatt gactggcaag atgtcccag aggtctcat	9840
	tcagcaataa aattctcacc ttcacccagg cccactgagt gtcagatttg catgcactag	9900
30	ttcacgtgtg taaaaaggag gatgcttctt tcctttgtat tctcacatac ctttaggaaa	9960
	gaacttagca cccttcccac acagccatcc caataactca tttcagtgac tcaacccttg	10020
35	actttataaa agtcttgggc agtatagagc agagattaag agtacagatg ctggagccag	10080
	accacctgag tgattagtga ctcagtttct cttagtagtt gtatgactca gtttcttcat	10140
	ctgtaaaatg gagggttttt taattagttt gtttttgaga aagggctctca ctctgtcacc	10200
40	caaagtgggag tgtagtggca aaatctcggc tcaactgcaac ttgcacttcc caggctcaag	10260
	cggctcctcc acctcaacat cctgagtagc tggaaccaca ggtacacacc accatacctc	10320
	gctaattttt tgtatttttg gtagagatgg ggtttcacat gttacacagg atgggtctcag	10380
45	actccggagc tcaagcaatc tgccacctc agccttccaa agtgctggga ttataagcat	10440
	gattacagga gttttaacag gtcataaga ttgttctgca gcccgagtga gttaatacat	10500
	gcaaagagtt taaagcagtg acttataaat gctaactact ctagaaatgt ttgctagtat	10560
50	tttttgttta actgcaatca ttcttgctgc aggtgaaaac tagtgttctg tactttatgc	10620
	ccattcatct ttaactgtaa taataaaaat aactgacatt tattgaaggc tatcagagac	10680
55	tgtaattagt gctttgcata attaatcata tttaatactc ttggattctt tcaggtagat	10740
	actattatta tccccatttt actacagtta aaaaaactac ctctcaactt gctcaagcat	10800

EP 3 620 521 A1

	acactctcac	acacacaaac	ataaactact	agcaaatagt	agaattgaga	tttggtccta	10860
	attatgtctt	tgctcactat	ccaataaata	tttattgaca	tgtacttctt	ggcagtctgt	10920
5	atgctggatg	ctggggatac	aaagatgttt	aaatttaagc	tccagtctct	gcttccaaag	10980
	gcctcccagg	ccaagttatc	cattcagaaa	gcatttttta	ctctttgcat	tccactgttt	11040
	ttcctaagtg	actaaaaaat	tacactttat	tcgtctgtgt	cctgctctgg	gatgatagtc	11100
10	tgactttcct	aacctgagcc	taacatccct	gacatcagga	aagactacac	catgtggaga	11160
	aggggtggtg	gttttgattg	ctgctgtctt	cagttagatg	gttaactttg	tgaagttgaa	11220
	aactgtggct	ctctggttga	ctgttagagt	tctggcactt	gtcactatgc	ctattattta	11280
15	acaaatgcat	gaatgcttca	gaatatggga	atattatctt	ctggaatagg	gaatcaagtt	11340
	atattatgta	accaggatt	agaagattct	tctgtgtgta	agaatttcat	aaacattaag	11400
	ctgtctagca	aaagcaaggg	cttgaaaaat	ctgtgagctc	ctcaccatat	agaaagcttt	11460
20	taacccatca	ttgaataaat	ccctataggg	gatttctacc	ctgagcaaaa	ggctggtctt	11520
	gattaattcc	caaactcata	tagctctgag	aaagtctatg	ctgttaacgt	tttcttgtct	11580
25	gctaccccat	catatgcaca	acaataaatg	caggcctagg	catgactgaa	ggctctctca	11640
	taattcttgg	ttgcatgaat	cagattatca	acagaaatgt	tgagacaaac	tatggggaag	11700
	cagggtatga	aagagctctg	aatgaaatgg	aaaccgcaat	gcttcctgcc	cattcagggc	11760
30	tccagcatgt	agaaatctgg	ggctttgtga	agactggctt	aaaatcagaa	gccccattgg	11820
	ataagagtag	ggaagaacct	agagcctacg	ctgagcaggt	ttccttcatg	tgacagggag	11880
	cctcctgccc	cgaacttcca	gggatccctc	cttaagtgtt	tcctgctgga	atctcctcac	11940
35	ttctatctgg	aaatggtttc	tccacagtcc	agcccctggc	tagttgaaag	agttacccat	12000
	gcagaggccc	tcctagcatc	cagagactag	tgcttagatt	cctactttca	gcgttgga	12060
	acctggatcc	acttgcccag	tgttcttcct	tagttcctac	cttcgacctt	gatcctcctt	12120
40	tatcttcctg	aacctgctg	agatgatcta	tgtggggaga	atggcttctt	tgagaaacat	12180
	cttcttcggt	agtggcctgc	ccctcattcc	cactttaata	tccagaatca	ctataagaag	12240
	aatataataa	gaggaataac	tcttattata	ggtaagggaa	aattaagagg	catacgtgat	12300
45	gggatgagta	agagaggaga	gggaaggatt	aatggacgat	aaaatctact	actatttggt	12360
	gagacctttt	atagtcta	caattttgct	attgttttcc	atcctcacgc	taactccata	12420
	aaaaaacact	attattatct	ttattttgcc	atgacaagac	tgagctcaga	agagtcaagc	12480
50	atttgcctaa	ggtcggacat	gtcagaggca	gtgccagacc	tatgtgagac	tctgcagcta	12540
	ctgctcatgg	gccctgtgct	gcactgatga	ggaggatcag	atggatgggg	caatgaagca	12600
55	aaggaatcat	tctgtggata	aaggagacag	ccatgaagaa	gtctatgact	gtaaatttgg	12660
	gagcaggagt	ctctaaggac	ttggatttca	aggaattttg	actcagcaaa	cacaagaccc	12720

EP 3 620 521 A1

	tcacggtgac	tttgcgagct	ggtgtgccag	atgtgtctat	cagaggttcc	agggaggggtg	12780
	gggtgggggtc	agggctggcc	accagctatc	agggcccaga	tgggttatag	gctggcaggc	12840
5	tcagataggt	ggttaggtca	ggttggtggt	gctgggtgga	gtccatgact	cccaggagcc	12900
	aggagagata	gaccatgagt	agagggcaga	catgggaaag	gtgggggagg	cacagcatag	12960
	cagcattttt	cattctacta	ctacatggga	ctgctccctt	atacccccag	ctagggggcaa	13020
10	gtgccttgac	tcctatgttt	tcaggatcat	catctataaa	gtaagagtaa	taattgtgtc	13080
	tatctcatag	ggttattatg	aggatcaaag	gagatgcaca	ctctctggac	cagtggccta	13140
	acagttcagg	acagagctat	gggcttcccta	tgtatgggtc	agtgggtctca	atgtagcagg	13200
15	caagttccag	aagatagcat	caaccactgt	tagagatata	ctgccagtct	cagagcctga	13260
	tgttaattta	gcaatgggct	gggaccctcc	tccagtagaa	ccttctaacc	agctgctgca	13320
20	gtcaaagtcg	aatgcagctg	gttagacttt	ttttaatgaa	agcttagctt	tcattaaaga	13380
	ttaagctcct	aagcagggca	cagatgaaat	tgtctaacag	caactttgcc	atctaaaaaa	13440
	atctgacttc	actggaaaca	tggaagccca	aggttctgaa	catgagaaat	ttttaggaat	13500
25	ctgcacagga	gttgagaggg	aaacaagatg	gtgaagggac	tagaaaccac	atgagagaca	13560
	cgaggaaata	gtgtagattt	aggctggagg	taaatgaaag	agaagtggga	attaatactt	13620
	actgaaatct	ttctatatgt	caggtgccat	tttatgatat	ttaataatct	cattacatat	13680
30	ggtaattctg	tgagatatgt	attattgaac	atactataat	taataactaat	gataagtaac	13740
	acctcttgag	tacttagtat	atgctagaat	caaatttaag	tttatcatat	gaggccgggc	13800
35	acggtggctc	atatatggga	ttacatgcct	gtaatcccag	cactttggga	ggccaaggca	13860
	attggatcac	ctgaggtcag	gagttccaga	ccagcctggc	caacatggtg	aaaccccttc	13920
	tctactaaaa	aatacaaaaa	atcagccagg	tgtggtggca	cgcgtctata	atcccagcta	13980
40	ctcaggaggc	tgaggcagga	gaatcacttg	aaccagggag	gtggaggttg	cagttagcta	14040
	agattgcacc	actgcactcc	agcctaggcg	acagagttag	actccatctc	aaaaaaaaaa	14100
	aaagaagttt	attatatgaa	ttaacttagt	tttactcaca	ccaatactca	gaagtagatt	14160
45	attacctcat	ttattgatga	ggagcccaat	gtacttgtag	tgtagatcaa	cttattgaaa	14220
	gcacaagcta	ataagtagac	aattagtaat	tagaagtcag	atggtctgag	ctctcctact	14280
	gtctacatta	catgagctct	tattaactgg	ggactcgaaa	atcaaagaca	tgaaataatt	14340
50	tgtccaagct	tacagaacca	ccaagtagta	aggctaggat	gtagaccag	ttctgctacc	14400
	tctgaagaca	gtgttttttc	cacagcaaaa	cacaaactca	gatattgtgg	atgcgagaaa	14460
55	ttagaagtag	atattcctgc	cctgtggccc	ttgcttctta	cttttacttc	ttgtcgattg	14520
	gaagttgtgg	tccaagccac	agttgcagac	catacttctt	caaccataat	tgcatttctt	14580

EP 3 620 521 A1

	caggaaagtt	tgagggagaa	aaaggtaaag	aaaaatttag	aaacaacttc	agaataaaga	14640
	gattttctct	tgggttacag	agattgtcat	atgacaaatt	ataagcagac	acttgagaaa	14700
5	actgaaggcc	catgcctgcc	caaattaccc	tttgacccct	tgggtcaagct	gcaacttttg	14760
	ttaaaggagg	tgtttatgtg	ttatagtgtt	cattttactct	tctgggtctaa	cccattggct	14820
	ccgtcttcat	cctgcagtga	cctcagtgcc	tcagaaacat	acatatgttt	gtctagttta	14880
10	agtttgtgtg	aaattctaac	tagcgtcaag	aactgagggc	cctaaactat	gctaggaata	14940
	gtgctgtggt	gctgtgatag	gtacacaaga	aatgagaaga	aactgcagat	tctctgcac	15000
	tccctttgcc	gggtctgaca	acaaagtttc	cccaaatttt	accaatgcaa	gccattttct	15060
15	catatgctaa	ctacttttaa	atcatttggg	gcttcacatt	gtctttctca	tctgtaaaaa	15120
	gaatggaaga	actcattcct	acagaactcc	ctatgtcttc	cctgatgggc	tagagttcct	15180
	ctttctcaaa	aattagccat	tattgtattt	ccttctaagc	caaagctcag	aggtcttgta	15240
20	ttgccagtg	acatgcacac	tgggtcaaaag	taggctaagt	agaagggtag	tttcacagga	15300
	acagagagca	aaagaggtgg	gtgaatgaga	gggtaagtga	gaaaagacaa	atgagaagtt	15360
25	acaacatgat	ggcttgttgt	ctaaatatct	cctaggggaat	tattgtgaga	ggtctgaata	15420
	gtgttgtaaa	ataagctgaa	tctgctgcca	acattaacag	tcaagaaata	cctccgaata	15480
	actgtacctc	caattattct	ttaaggtagc	atgcaactgt	aatagttgca	tgtatatatt	15540
30	tatcataata	ctgtaacaga	aaacacttac	tgaatatata	ctgtgtccct	agttctttac	15600
	acaataaaact	aatctcatcc	tcataattct	attagctaata	acatattatc	atcctatatt	15660
	tcagagactt	caagaagtta	agcaacttgc	tcaagatcat	ctaagaagta	ggtggtattt	15720
35	ctgggctcat	ttggcccttc	ctaactcttc	atggcaacat	ggctgcctaa	agtggtgatt	15780
	gccttaattc	atcagggatg	ggctcatact	cactgcagac	cttaactggc	atcctctttt	15840
	cttatgtgat	ctgcctgacc	ctagtagact	tatgaaattt	ctgatgagaa	aggagagagg	15900
40	agaaaggcag	agctgactgt	gatgagtgat	gaagggtgct	tctcatctgg	gtaccagtgg	15960
	ggcctctaag	actaagtcac	tctgtctcac	tgtgtcttag	ccagttcctt	acagcttgcc	16020
	ctgatgggag	atagagaatg	ggtatcctcc	aacaaaaaaa	taaattttca	tttctcaagg	16080
45	tccaacttat	gttttcttaa	tttttaaaaa	aatcttgacc	attctccact	ctctaaaata	16140
	atccacagtg	agagaaacat	tcttttcccc	catcccataa	atacctctat	taaatatgga	16200
	aaatctgggc	atggtgtctc	acacctgtaa	tcccagcact	ttgggaggct	gaggtgggtg	16260
50	gactgcttgg	agctcaggag	ttcaagacca	tcttgacaaa	catggtgata	ccctgcctct	16320
	acaaaaagta	caaaaattag	cctggcatgg	tgggtgtgcac	ctgtaatccc	agctattagg	16380
55	gtggctgagg	caggagaatt	gcttgaaccc	gggaggcgga	ggttgcagtg	agctgagatc	16440
	gtgccactgc	actccagcct	gggggacaga	gcacattata	attaactgtt	attttttact	16500

EP 3 620 521 A1

	tggactcttg	tggggaataa	gatacatggt	ttattcttat	ttatgattca	agcactgaaa	16560
	atagtgttta	gcatccagca	ggtgcttcaa	aaccatttgc	tgaatgatta	ctatactttt	16620
5	tacaagctca	gtccctcta	tcccttccag	catcctcatc	tctgattaaa	taagcttcag	16680
	tttttcctta	gttcctgtta	catttctgtg	tgtctccatt	agtgacctcc	catagtccaa	16740
	gcatgagcag	ttctggccag	gcccctgtcg	gggtcagtcg	cccccccccg	ccttctgggt	16800
10	ctgtgtaacc	ttctaagcaa	accttctggc	tcaagcacag	caatgctgag	tcatgatgag	16860
	tcatgctgag	gcttaggggtg	tgtgcccaga	tgttctcagc	ctagagtgat	gactcctatc	16920
	tgggtcccca	gcaggatgct	tacagggcag	atggcaaaaa	aaaggagaag	ctgaccacct	16980
15	gactaaaact	ccacctcaaa	cggcatcata	aagaaaatgg	atgcctgaga	cagaatgtga	17040
	catattctag	aatatattat	ttcctgaata	tatatatata	tatacacata	tacgtatata	17100
20	tatatatata	tatatatttg	ttgttatcaa	ttgccataga	atgattagtt	attgtgaatc	17160
	aaatatttat	cttgcagggtg	gcctctatac	ctagaagcgg	cagaatcagg	ctttattaat	17220
	acatgtgtat	agattttttag	gatctatata	catgtattaa	tatgaaacaa	ggatatggaa	17280
25	gaggaaggca	tgaaaacagg	aaaagaaaac	aaaccttggt	tgccatttta	aggcaccctt	17340
	ggacagctag	gtggcaaaaag	gcctgtgctg	ttagaggaca	catgctcaca	tacgggggtca	17400
	gatctgactt	ggggtgctac	tgggaagctc	tcatcttaag	gatacatctc	aggccagtct	17460
30	tgggtgcatta	ggaagatgta	ggcaactctg	atcctgagag	gaaagaaaca	ttcctccagg	17520
	agagctaaaa	gggttcacct	gtgtgggtaa	ctgtgaagga	ctacaagagg	atgaaaaaca	17580
35	atgacagaca	gacataatgc	ttgtgggaga	aaaaacagga	ggtcaagggg	atagagaagg	17640
	cttccagaag	aatggctttg	aagctggctt	ctgtaggagt	tcacagtggc	aaagatgttt	17700
	cagaaatgtg	acatgactta	aggaactata	caaaaaggaa	caaatttaag	gagaggcaga	17760
40	taaattagtt	caacagacat	gcaaggaatt	ttcagatgaa	tgttatgtct	ccactgagct	17820
	tcttgagggt	agcagctgtg	agggttttgc	aggcccagga	cccattacag	gacctcacgt	17880
	atacttgaca	ctgttttttg	tattcatttg	tgaatgaatg	acctcttgtc	agtctactcg	17940
45	gtttcgctgt	gaatgaatga	tgtcttgtca	gcctacttgg	tttcgctaag	agcacagaga	18000
	gaagatttag	tgatgctatg	taaaaacttc	ctttttgggt	caagtgtatg	tttgtgatag	18060
	aaatgaagac	aggctacatg	atgcatatct	aacataaaca	caaacattaa	gaaaggaaat	18120
50	caacctgaag	agtattttata	cagataacaa	aatacagaga	gtgagttaaa	tgtgtaataa	18180
	ctgtggcaca	ggctggaata	tgagccattt	aaatcacaaa	ttaattagaa	aaaaaacagt	18240
55	ggggaaaaaa	ttccatggat	gggtctagaa	agactagcat	tgttttaggt	tgagtggcag	18300
	tgtttaaagg	gtgatatcag	actaaacttg	aaatatgtgg	ctaaataact	agaatactct	18360

EP 3 620 521 A1

	ttatTTTTTtc	gtatcatgaa	tagcagatat	agcttgatgg	cccatgctt	ggtttaacat	18420
	ccttgctgtt	cctgacatga	aatccttaat	ttttgacaaa	ggggctattc	attttcattt	18480
5	tatattgggc	ctagaaatta	tgtagatgg	cctgaggaaa	agtttatagc	ttgtctattt	18540
	ctctctctaa	catagttgtc	agcacaatgc	ctaggctata	ggaagtactc	aaagcttggt	18600
	aaattgaatt	ctatccttct	tattcaattc	tacacatgga	ggaaaaactc	atcagggatg	18660
10	gaggcacgcc	tctaaggaag	gcaggtgtgg	ctctgcagtg	tgattgggta	cttgaggac	18720
	gaagggtggg	gtgggagtgg	ctaacccttc	attcctagt	cagaggtcac	agcctaaaca	18780
	tcaaatctct	tgagggtcgg	tggctcactc	ctgtaatcac	agcagtttgg	gacgccaaag	18840
15	tgggcagatc	acttgagggtc	aggagtgtga	caccagccca	gccaacatag	tgaaacctgg	18900
	tctctgctta	aaaatataaa	aattagctgg	acgtggtgac	gggagcctgt	aatccaacta	18960
	cttgggaggc	tgaggcagga	gaatcgcttg	aaccggggag	gtggagtgtg	cactgagcag	19020
20	agatcatgcc	attgcactcc	agcctccaga	gcgagactct	gtctaaagaa	aaacgaaaac	19080
	aaacaaacaa	acaaacaaac	aaaacccttc	aaattccctg	accgaacaga	attctgtctg	19140
25	attgttctct	gacttatcta	ccattttccc	tccttaaaga	aactgtgaac	ttccttcagc	19200
	tagaggggcc	tggctcagaa	gcctctggtc	agcatccaag	aaatacttga	tgtcactttg	19260
	gctaaaggta	tgatgtgtag	acaagctcca	gagatggttt	ctcatttcca	tatccaccca	19320
30	cccagctttc	caatttttaa	gccaatcttg	aggtagagac	tgtgatgaac	aaacaccttg	19380
	acaaaattca	acccaaagac	tcactttgcc	tagcttcaaa	atccttactc	tgacatatac	19440
	tcacagccag	aaattagcat	gcactagagt	gtgcatgagt	gcaacacaca	cacacaccaa	19500
35	ttccatattc	tctgtcagaa	aatcctgttg	gtttttcgtg	aaaggatgtt	ttcagaggct	19560
	gaccccttgc	cttcacctcc	aatgctacca	ctctgggtcta	agtcactgtc	accaccacct	19620
	aaattatagc	tgttgactca	taacaatctt	cctgcttcta	ccactgcccc	actacaattt	19680
40	cttcccaata	tactatccaa	attagtcttt	tcaaaatgta	agtcataatat	ggtcacctct	19740
	ttgttcaaag	tcttctgata	gtttcctata	tcatttataa	taaaaccaa	tccttacaat	19800
	tctctacaat	agttgttcat	gcatatatta	tgtttattac	agatacatat	atatagctct	19860
45	catataaata	aatatatata	tttatgtgta	tgtgtgtaga	gtgttttttc	ttacaactct	19920
	atgatgtagg	tattattagt	gtcccaaatt	ttataattta	ggacttctat	gatctcatct	19980
	tttattctcc	ccttcaccga	atctcatcct	acattggcct	tattgatatt	ccttgaaaat	20040
50	tctaagcatc	ttacatcttt	agggtattta	catttgccat	tccttatgcc	ctaaatattt	20100
	aatcatagtt	tcatataaat	gggttcctca	tcacttatgg	gtactctctc	aggtgttaac	20160
55	tttatagtga	ggactttcct	gccatactac	ttaaagtagc	gatacccttt	caccctgtcc	20220
	taatcacact	ctggccttca	tttcagtttt	ttttttttct	ccatagcacc	taatctcatt	20280

EP 3 620 521 A1

	ggtatataac atgttttcatt tgcttatttta atgtcaagct ctttccacta tcaagtccat	20340
	gaaaacagga actttattcc tctattctgt ttttgtgctg tattcttagc aattttacaa	20400
5	ttttgaatga atgaatgagc agtcaaacac atatacaact ataattaaaa ggatgtatgc	20460
	tgacacatcc actgctatgc acacacaaag aaatcagtggt agtagagctg gaagtgtctaa	20520
	gcctgcatag agctagtttag ccctccgcag gcagagcctt gatgggatta ctgagttcta	20580
10	gaattggact catttggtttt gtaggctgag atttgctctt gaaaacttgt tctgacccaa	20640
	ataaaaggct caaaagatga atatcgaaac cagggtgttt tttaactgg aattttataac	20700
	tagagcactc atgtttatgt aagcaattaa ttgtttcatc agtcaggtaa aagtaaagaa	20760
15	aaactgtgcc aaggcaggta gcctaatagca atatgccact aaagtaaaca ttatttcata	20820
	ggtgtcagat atggcttatt catccatctt catgggaagg atggccttgg cctggacatc	20880
20	agtgttatgt gaggttcaaa acacctctag gctataaggc aacagagctc cttttttttt	20940
	tttctgtgct ttccctggctg tccaaatctc taatgataag catacttcta ttcaatgaga	21000
	atattctgta agattatagt taagaattgt gggagccatt ccgtctctta tagttaaatt	21060
25	tgagcttctt ttatgatcac tgttttttta atatgcttta agttctgggg tacatgtgcc	21120
	atggtgggtt gctgcaccca tcaaccctgc atctacatta ggtatttctc ctaatgctat	21180
	ccttccccta gccccccacc cccaacaggc ccagtggtgt gatgttcccc tccctgtgtc	21240
30	catggatcac tgggtttttt ttgttttttt ttttttttta aagtctcagt taaatttttg	21300
	gaatgtaatt tattttcctg gtatccctagg acttgcaagt tatctgggtca ctttagccct	21360
35	cacgttttga tgataatcac atatttgtaa acacaacaca cacacacaca cacacacaca	21420
	tatatatata tataaaacat atatatacat aaacacacat aacatattta tcggggcattt	21480
	ctgagcaact aatcatgcag gactctcaaa cactaaccta tagccttttc tatgtatcta	21540
40	cttgtgtaga aaccaagcgt ggggactgag aaggcaatag caggagcatt ctgactctca	21600
	ctgccttttag ctaggcccct ccctcatcac agctcagcat agtcctgagc tcttatctat	21660
	atccacacac agtttctgac gctgcccagc tatcaccatc ccaagtctaa agaaaaaat	21720
45	aatgggtttg cccatctctg ttgattagaa aacaaaacaa aataaaataa gccctaagc	21780
	tcccagaaaa catgactaaa ccagcaagaa gaagaaaata caataggtat atgaggagac	21840
	tggtgacact agtgtctgaa tgaggcttga gtacagaaaa gaggtcttag cagcatagtg	21900
50	gttttagagga gatgtttctt tccttcacag atgccttagc ctcaataagc ttgcggttgt	21960
	ggaagtttac tttcagaaca aactcctgtg gggctagaat tattgatggc taaaagaagc	22020
55	ccgggggagg gaaaaatcat tcagcatcct cacccttagt gacacaaaac agaggggggc	22080
	tgggtttcca tatttcctca tgatggatga tctcgttaat gaaggtgggtc tgacgagatc	22140

EP 3 620 521 A1

	attgcttctt	ccatttaagc	cttgctcact	tgccaatcct	cagttttaac	cttctccaga	22200
	gaaatacaca	ttttttattc	aggaaacata	ctatgttata	gtttcaatac	taaataatca	22260
5	aagtactgaa	gatagcatgc	ataggcaaga	aaaagtcctt	agctttatgt	tgctgttggt	22320
	tcagaattta	aaaaagatca	ccaagtcaag	gactttctcag	ttctagcact	agaggtggaa	22380
	tcttagcata	taatcagagg	tttttcaaaa	tttctagaca	taagattcaa	agccctgcac	22440
10	ttaaaatagt	ctcatttgaa	ttaactcttt	atataaattg	aaagcacatt	ctgaactact	22500
	tcagagtatt	gttttatttc	tatgttctta	gttcataaat	acattaggca	atgcaattta	22560
	attaaaaaaa	cccaagaatt	tcttagaatt	ttaatcatga	aaataaatga	agggcatcttt	22620
15	acttactcaa	gggtcccaaaa	ggtcaaagaa	accaggaaag	taaagctata	tttcagcgga	22680
	aaatgggata	tttatgagtt	ttctaagttg	acagactcaa	gttttaacct	tcagtgccca	22740
	tcatgtagga	aagtgtggca	taactggctg	attctggctt	tctactcctt	tttccatta	22800
20	aagatccctc	ctgcttaatt	aacattcaca	agtaactctg	gttgtagctt	agggcacagt	22860
	gctcccgagg	tcagtcacac	aataggatgt	ctgtgctcca	agttgccaga	gagagagatt	22920
	actcttgaga	atgagcctca	gccctggctc	aaactcacct	gcaaacttcg	tgagagatga	22980
25	ggcagaggta	cactacgaaa	gcaacagtta	gaagctaaat	gatgagaaca	catggactca	23040
	tagagggaaa	caacgcatac	tggggcctat	cagaggggtg	aggggtgagag	aaggagagga	23100
	tcaggaaaaa	tcactaatgg	atgctaagcg	taatacctga	gtgatgagat	catctatata	23160
30	acaaaccccc	ttgacattca	tttatctatg	taacaaacct	gcacatcctg	tacatgtacc	23220
	cctgaactta	aaataaaagt	tgaaaacaag	aaagcaacag	tttgaacact	tggtatggtc	23280
35	tattctctca	ttcttttaca	ttacactaga	aaatagccac	aggcttcctg	caaggcagcc	23340
	acagaattta	tgacttggtg	tatccaagtc	attcctggat	aatgcaaaat	ctaacacaaa	23400
	atctagtaga	atcatttgct	tacatctatt	tttgttctga	gaatatagat	ttagatacat	23460
40	aatggaagca	gaataattta	aaatctggct	aatttagaat	cctaagcagc	tcttttctta	23520
	tcagtgggtt	acaagccttg	tttatatttt	tcctattttta	aaaataaaaa	taaagtaagt	23580
	tatttggtgt	aaagaatatt	cattaaagta	tttatttctt	agataatacc	atgaaaaaca	23640
45	ttcagtgaag	tgaagggcct	actttactta	acaagaatct	aatttatata	atttttcata	23700
	ctaatagcac	ctaagaacag	tacaatatct	gactcttcag	gttaaacata	tgtcataaat	23760
	tagccagaaa	gattttaagaa	aatattggat	gtttccttgt	ttaaattagg	catcttacag	23820
50	tttttagaat	cctgcataga	acttaagaaa	ttacaaatgc	taaagcaaac	ccaacagggc	23880
	aggaattaat	cttcatcgaa	tttgggtgtt	tctttctaaa	agtcctttat	acttaaagt	23940
55	cttaagacac	acatagattt	tattttacta	attttaatta	tatagacaat	aatgaatat	24000
	tcttactgat	tactttttct	gactgtctaa	tctttctgat	ctatcctgga	tggccataac	24060

EP 3 620 521 A1

	acttatctct	ctgaactttg	ggctttttaat	ataggaaaga	aaagcaataa	tccatttttc	24120
	atgggtatctc	atatgataaa	caaataaaat	gcttaaaaaat	gagcaggtga	agcaatttat	24180
5	cttgaaccaa	caagcatcga	agcaataatg	agactgcccc	cagcctacct	gacttctgag	24240
	tcaggattta	taagccttgt	tactgagaca	caaacctggg	cctttcaatg	ctataacctt	24300
	tcttgaagct	cctccctacc	accttttagcc	ataaggaaac	atggaatggg	tcagatccct	24360
10	ggatgcaagc	caggtctgga	accataggca	gtaaggagag	aagaaaatgt	gggctctgca	24420
	actggctccg	agggagcagg	agaggatcaa	ccccatactc	tgaatctaag	agaagactgg	24480
	tgtccatact	ctgaatggga	agaatgatgg	gattacccat	agggcttggt	ttaggggagaa	24540
15	acctgttctc	caaactcttg	gccttgagat	acctggctct	tattccttgg	actttggcaa	24600
	tgtctgaccc	tcacattcaa	gttctgagga	agggccactg	ccttcatact	gtggatctgt	24660
20	agcaaattcc	ccctgaaaac	ccagagctgt	atcttaattg	gttaaaaaaa	attatattat	24720
	ctcaacgact	gttcttctct	gagtagccaa	gctcagcttg	gttcaagcta	caagcagctg	24780
	agctgctttt	tgtctagtca	ttgttctttt	atttcagtgg	atcaaatacg	ttctttccaa	24840
25	acctaggatc	ttgtcttcct	aggctatata	ttttgtccca	ggaagtctta	atctggggtc	24900
	cacagaacac	tagggggctg	gtgaagttta	tagaaaaaaa	atctgtattt	ttacttacat	24960
	gtaactgaaa	tttagcattt	tcttctactt	tgaatgcaaa	ggacaaacta	gaatgacatc	25020
30	atcagtacct	attgcatagt	tataaagaga	aaccacagat	attttcatac	tacaccatag	25080
	gtattgcaga	tcttttttgt	tttgtttttg	tttgagatgg	agtttcgctc	ttattgcccc	25140
35	ggctggagtg	cagtggcatg	atttcggctc	actgcaacct	ccccctcctg	cattcaagca	25200
	attctcctgc	cttggcctcc	tgagtagctg	gggattacag	gcacctgcc	ccatgccagt	25260
	ctaatttttg	tatttttagt	agagatgggg	tttcgccatg	ttggccaggc	tggtcttgaa	25320
40	ctcctgacct	cagatgatct	gcccgccttg	gcctcctgaa	gtgctgggat	tataggtgtg	25380
	agccaccacg	cctggcccat	tgcagatatt	tttaattcac	atttatctgc	atcactactt	25440
	ggatcttaag	gtagctgtag	acccaatcct	agatctaata	ccttcataaa	gaagcaaata	25500
45	taataaatac	tataaccaca	atgtaatgtt	tgatgtctga	taatgatatt	tcagtgtaat	25560
	taaacttagc	actcctatgt	atattatttg	atgcaataaa	aacatatttt	tttagcactt	25620
	acagtctgcc	aaactggcct	gtgacacaaa	aaaagttag	gaattcctgg	ttttgtctgt	25680
50	gttagccaat	ggttagaata	tatgctcaga	aagataccat	tggttaatag	ctaaaagaaa	25740
	atggagtaga	aattcagtgg	cctggaataa	taacaatttg	ggcagtcatt	aagtcaggtg	25800
55	aagacttctg	gaatcatggg	agaaaagcaa	gggagacatt	cttacttgcc	acaagtgttt	25860
	tttttttttt	ttttttttat	cacaaacata	agaaaatata	ataaataaca	aagtcaggtt	25920

EP 3 620 521 A1

	atagaagaga gaaacgctct tagtaaactt ggaatatgga atccccaaag gcacttgact	25980
	tgggagacag gagccatact gctaagtga aaagacgaag aacctctagg gcctgaacat	26040
5	acaggaaatt gtaggaacag aaattcctag atctggtggg gcaaggggag ccataggaga	26100
	aagaaatggt agaaatggat ggagacggag gcagaggtgg gcagatcatg aggtcaagag	26160
	atcgagacca tcctggcaaa catggtgaaa tcccgtctct actaaaaata aaaaaattag	26220
10	ctgggcatgg tggcatgctc ctgtagtccc agctgctcgg gaggctgagg caggagaatc	26280
	gtttgaaccc aggaggcgaa ggttgcaagt agctgagata gtgccattgc actccagtct	26340
	ggcaacagag tgagactccg tctcaaaaaa aaaaaaaaaa gaaagaaaga aaagaaaaag	26400
15	aaaaaagaaa aaataaatgg atgtagaaca agccagaagg aggaactggg ctggggcaat	26460
	gagattatgg tgatgtaagg gacttttata gaattaacaa tgctggaatt tgtggaactc	26520
	tgcttctatt attcccccaa tcattacttc tgtcacattg atagttaa attttctgtg	26580
20	aattttattcc ttgattctaa aatatgagga taatgacaat ggtattataa gggcagatta	26640
	agtgatatag catgagcaat attcttcagg cacatggatc gaattgaata cactgtaa at	26700
	cccaacttcc agtttcagct ctaccaagta aagagctagc aagtcatcaa aatggggaca	26760
25	tacagaaaaa aaaaaggaca ctagaggaat aatataccct gactcctagc ctgattaata	26820
	tatcgattca cttttttctc tgtttgatga caaattctgg ctttaaataa ttttaggatt	26880
30	ttaggcttct cagctccctt ccagtgaga agtataagca ggacagacag gcaagcaaga	26940
	agagagcccc aggcaatact cacaaagtag ccaatgtccc ctgtggtcat agagaaatga	27000
	aaagagagag gattctctgg aagcactgga tgtaatcttt tctgtctgtc ctctctaggg	27060
35	aatcacccca aggtactgta ctttgggatt aaggcttttag tcccactgtg gactacttgc	27120
	tattctgttc agtttctaga aggaactatg tacgggtttt gtctccctag agaaactaag	27180
	gtacagaagt tttgtttaca atgcactcct taagagagct agaactgggt gagattctgt	27240
40	tttaacagct ttattttctt ttccttgccc ctgtttttgt cactgtcacc acctttaagg	27300
	caaagtgtta atgcgctttg gctgaaactt tttttcctat tttgagattt gtccttttat	27360
	atgaggcttt cttggaaaag gagaatggga gagatggata tcatttttga agatgatgaa	27420
45	gagggtaaaa aaggggacaa atggaaattt gtgttgacaga tagatgagga gccaacaaaa	27480
	aagagcctca ggatccagca cacattatca caaacttagt gtccatccat cactgctgac	27540
	cctctccgga cctgactcca cccctgaggg acacaggtca gccttgacca atgactttta	27600
50	agtaccatgg agaacagggg gccagaactt cggcagtaaa gaataaaagg ccagacagag	27660
	aggcagcagc acatatctgc ttccgacaca gctgcaatca ctagcaagct ctcaggcctg	27720
55	gcatcatggt gcattttact gctgaggaga aggctgccgt cactagcctg tggagcaaga	27780
	tgaatgtgga agaggctgga ggtgaagcct tgggcaggta agcattgggt ctcaatgcat	27840

EP 3 620 521 A1

	gggaatgaag ggtgaatatt accctagcaa gttgattggg aaagtcctca agatTTTTTg	27900
	catctctaata tttgtatctg atatggtgtc atttcataga ctctctggtt tttaccctg	27960
5	gaccagaga ttttttgaca gctttggaaa cctgtcgtct ccctctgcca tcctgggcaa	28020
	ccccaaaggtc aaggcccatg gcaagaaggt gctgacttcc tttggagatg ctattaaaaa	28080
	catggacaac ctcaagcccg cctttgctaa gctgagttag ctgcactgtg acaagctgca	28140
10	tgtggatcct gagaacttca aggtgagttc aggtgctggg gatgtgattt tttggcttta	28200
	tattttgaca ttaattgaag ctcataatct tattggaaag accaacaag atctcagaaa	28260
	tcattgggtcg agcttgatgt tagaacagca gacttctagt gagcataacc aaaacttaca	28320
15	tgattcagaa ctagtgacag taaaggacta ctaacagcct gaattggctt aacttttcag	28380
	gaaatcttgc cagaacttga tgtgtttatc ccagagaatt gtattataga attgtagact	28440
20	tgtgaaagaa gaatgaaatt tggcttttgg tagatgaaag tccatttcaa ggaaatagaa	28500
	atgccttatt ttatgtgggt catgataatt gaggtttaga aagagatttt tgcaaaaaaa	28560
	ataaaagatt tgctcaaaga aaaataagac acattttcta aaatatgtta aatttcccat	28620
25	cagtattgtg accaagtga ggcttgtttc cgaatttgtt ggggatttta aactcccgct	28680
	gagaactctt gcagcactca cattctacat ttacaaaaat tagacaattg cttaaagaaa	28740
	aacagggaga gagggaaacc aataatactg gtaaaatggg gaagggggtg aggggtgtagg	28800
30	taggtagaat gttgaatgta gggctcatag aataaaattg aacctagct catctgaatt	28860
	ttttgggtgg gcacaaacct tggaaacagtt tgaggtcagg gttgtctagg aatgtaggta	28920
35	taaagccgtt tttgtttgtt tgtttgtttt ttcatcaagt tgttttcgga aacttctact	28980
	caacatgcct gtgtgttatt ttgtcttttg cctaacagct cctgggtaac gtgatggtga	29040
	ttattctggc tactcacttt ggcaaggagt tcaccctga agtgcaggct gcctggcaga	29100
40	agctggtgtc tgctgtcgcc attgccctgg ccataagta ccactgagtt ctcttccagt	29160
	ttgcagggtg tcctgtgacc ctgacaccct ccttctgcac atggggactg ggcttggcct	29220
	tgagagaaag ccttctgttt aataaagtac attttcttca gtaatcaaaa attgcaattt	29280
45	tatcttctcc atcttttact cttgtgttaa aaggaaaaag tgttcatggg ctgagggatg	29340
	gagagaaaca taggaagaac caagagcttc cttaagaaat gtatgggggc ttgtaaaatt	29400
	aatgtggatg ttatgggaga attccaggat tccaaggagg atgatatgat ggagaaaaat	29460
50	ctttatcggg gtgggaaaaat ggttaattaa gtggacagag actcctaggc agtttttact	29520
	gcaccgggga aagaaggagc tgttagtggt acctgagaaa gcagatttgt ggtacatgtc	29580
55	acttttcatt aaaaacaaaa acaaaacaaa acaaaacttc atagatatcc aagatatagg	29640
	ctagaattac tattttaatt tactcttatt tacattttga agtagctagc ttgtcacatg	29700

EP 3 620 521 A1

	ttttatgaaa ttgatttggg gataagatga gtgtgtatca acaatagcct gctctttcca	29760
	tgaaggattc cattatttca tgggttagct gaagctaaga cacatgatat cattgtgcat	29820
5	tatcttctga tagaatgtaa catgcactaa aataaagtta gagttaggac ctgagtggga	29880
	aagtttttgg agagtgtgat gaagactttc cgtgggagat agaatactaa taaaggctta	29940
	aattctaaaa ccagcaagct agggcttcgt gacttgcacg aaactggctc tctggaagta	30000
10	gaagggagag taagacatac gtagaggact aggaaagacc agatagtaca gggcctggct	30060
	acaaaaatac aagctttttac tatgctattg caatactaaa cgataagcat taggatgtta	30120
	agtgactcag gaaataagat tttgggaaaa agtaatctgc ttatgtgcac aaaatggatt	30180
15	caagtttgca gataaaataa aatatggatg atgattcaag gggacagata caatggttca	30240
	aacccaagag gagcagtgag tctgtggaat ttgaaggatg gacaaagggtg gggtgagaaa	30300
	gacatagtat tgcactgact gtgggagatg agaaggaaga aggaggtgat aaatgactga	30360
20	aagctcccag actggtgaag ataacaggag gaaaccatgc actgacctgg tgactctcat	30420
	gtgtgaagggt tagagggata ttaacagatt tacttttttag gaagtgctag attggtcagg	30480
	gagttttgac cttcaggtct tgtgtctttc atatcaagga acctttgcat tttccaagtt	30540
25	agagtgccat attttggcaa atataacttt attagtaatt ttatagtgc ctcacattga	30600
	tcagactttt tcctgtgaat tacttttgaa tttggctgta tatatccaga atatgggaga	30660
	gagacaaata attattgtag ttgcaggcta tcaacaatac tgggtctctct gagccttata	30720
30	acctttcaat atgcccataa acagagtaaa cagggtattat tcatggcact aaatattttc	30780
	acctagtacg tcaacaaatg ggagcaatgt gcattttttg atacatattt ttatatattt	30840
35	atgggggtaca tgtgatactt acatgcctag aacatgtgat gattaagtct agatatttag	30900
	gatatccatt gctttgagca tttatcattt ctatgtattg agaaaatttc aaatcctcat	30960
	ttctagccat tttgaaatat ataataaata gtaattaact atagtcaccc tactcaaata	31020
40	tcaaacatta tggcttaatc cttctatcca actgtgtttg tacctattaa ccaacatctc	31080
	ttaaatcccc tcccatcac actcacactt tttccagcct ctgataacta tcattctact	31140
	ctctaccacc atgagacca ctttttttagc tcccacagat gaataaaaaac atgtgatatt	31200
45	tgactttctg tatctggctt attttattat ctatctcttt ggcataccaa gagtttggtt	31260
	ttgttctgct tcagggcttt caattaacat aatgacctct ggttccatcc atgttgctac	31320
	aaatgacaag atttcattct ttttcatggc aaaatagtag tgtgcaaaaa tacaattttt	31380
50	taatccgttc atctgttgat agacacttag gttgatccca aaccttaact attgtgaata	31440
	gtgcttcaat aaacatgagt gtaatgtgtc cattggatat actgatttcc tttcttttgg	31500
55	ataaataaacc actagtgaga ttgctggatt gtatgatagt tctgttttta gtttactgag	31560
	aaatcttcat actgttttcc ataatggttg tactatttta cattcccacc aacagtgtgt	31620

EP 3 620 521 A1

	aagaaagagt	tcccttttct	ccatatcctc	acaaggatct	gttatttttt	gtcttttttg	31680
	ttaatagccg	ttttaactag	agtaagtaga	tatctcattg	tagttttgat	ttgcatttcc	31740
5	ctgatcatta	gtgatgttga	gaattttttc	atatgtttgt	tggtcatttg	tatatctttt	31800
	tctgagaatt	gtctgttcat	gtccttagcc	tactttttat	tgggattggt	tggtattttc	31860
	ttgataatct	atttgtgttc	atttttagagc	ctggatatta	ttcttttgtc	agatgtatag	31920
10	attgtgaaga	ttttctccca	ctctgtgggt	tgtctgttta	ttctgcagac	tcttcctttt	31980
	gccatgcaaa	agctcttttag	tttaatttag	tcccagatat	tttctttggt	tttatgtatt	32040
	tgcatttgtg	ttcttgggtca	tgaaatcctt	tcctaagcca	atgtgtagaa	gggtttttcc	32100
15	gatgttattt	tctagaattg	ttacagtttc	agggttaga	tttaagtcct	tgatccatct	32160
	tgagttgatt	tttgatataag	gtgagagatg	aagatccagt	ttcattctcc	tacatgtagc	32220
20	ttgccagcta	tccccgcacc	atttgttgaa	tagggtgccc	tttccccact	ttatgttttt	32280
	gtttgctttg	tcaaagatca	gttggatgta	agtatttgag	tttatttctg	ggttctctat	32340
	tctgttccat	tggtcgatgt	gcctatttgt	acaccagcat	catgctgttt	tggtgactat	32400
25	ggccttattg	tatagtttga	aatgaggtaa	tgtaatgcct	tcagatttgt	tctttttttt	32460
	agacttgctt	gtttattggg	ctcttttttg	gttccataag	aatttttagga	ttgttttttc	32520
	tagttctgtg	aagactaatg	gtgggtattt	gatgggaatt	gcaatgaatt	tgtagggtgc	32580
30	ttctggcatt	atggccattt	tcacaatatt	gattctaccc	atctatgaga	atggcatgtg	32640
	tttccatttg	tttgtgtctt	atatgattac	tttcagccgt	gttttgtagt	tttccttgta	32700
35	gatgtctttc	acctccttgg	ttaggtatat	attcctaagt	ttttgttttg	ttttgttttg	32760
	ttttttgcag	ctattgtaaa	aggggttgag	ttcttgattt	tattctcagc	ttggtcattg	32820
	ctggtatgta	agaaagcaac	tcattgggtg	acgttaattt	tgtatccaga	aactttgctg	32880
40	aattatttta	tcagttctag	gggggttttg	aggagtcttt	agagttttct	acatacacia	32940
	tcatatcatc	agcaaacagt	gacagtttga	ctttctcttt	aacaatttgg	atgtgcttta	33000
	cttgtttctc	ttgtctgatt	gctcttgcta	ggacttccag	taatatgtta	aagagaagtg	33060
45	gtgagagtgg	gtatccttgt	ctcattccag	ttttcagaca	gaatgctttt	aactttttcc	33120
	cattcaatat	aatgttggct	gtgtgtttac	catagctggc	ttttattaca	ttgaggtatg	33180
	tcctttgtaa	accgattttg	ctgagtttta	gtcataaagt	gatgttgaat	tttgttgaat	33240
50	gcagtttctg	tggctattga	gataatcaca	tgatttttgt	ttccaattct	ctttatgttg	33300
	tgtatcacac	ttattgactt	gcgtatgtta	aaccatccgt	gcacccctcg	catgaaaccc	33360
55	acttgatcat	gggttttgat	atgctgtcgg	atgctattag	ctagtatttt	gtcaaggatg	33420
	ttggcatcta	tgttcatcag	ggatattgat	ctgtagtggt	tttttttttt	ggttatgttc	33480

EP 3 620 521 A1

	tttcccagtt ttggtattaa ggtgatactg gcttcataga atgatttagg gaggattctc	33540
	tctttctcta tctttagtaa tactgtcaat aggattggta tcaattcttc tttgaatgtc	33600
5	tggtagaatt cagctgtgaa tctatctggt cctggacttt tttgttggtg gtaaattttt	33660
	attatcattt cagtcttgct gcttattact ggtctgttca gggatatctaa ttcttcctga	33720
	cttaagctag agccctgtat ctttccagga attcgaacgt ctcccttagg ttttctagtt	33780
10	tatgcatgta aagggtgttca tagtagcctt gaataatctt ttgtatttct gtggtatcag	33840
	taatagtatc tcctgttttg tttctaattg agtttatttg cacttctctc ctcttttctt	33900
	ggttaatctt gctaattggtc tatcagtttt atttatcttt tcaaagaacc agctttttat	33960
15	ttcatttagc ttttgtattt ttttgcagtt gttttaattt catttagttc tcctcttata	34020
	ttagttattc cctttctttt gctgggtttt ggttctgttt gtttttgttt ctctagtttc	34080
	ttgtggtgtg accttatatt gtctgtctgt cctctttcag actctttgac atcgacattt	34140
20	agggtctgta actttccttt tagcaccatc tttgctgtat cctagagggt ttgatagggt	34200
	gtgtcactat tgtcggtcag ttcaagtaat tttgttggtc ttattatact ttaagttctg	34260
25	ggatacatgt gcagaatgtg caggtttggt acatagggtat agatgtgcca tgggtggttg	34320
	ctgcacccat caacctgtca tctacattag gtatttcttt taatgttatc cctctcctaa	34380
	ccccctcacc ccccgacagg ccctggtgtg tgatgttccc ctccctgtgt ccatgtgttc	34440
30	tcattgttca actcccactt atgagtgaga acgtgtggtg tttggtttct ctgttcctgt	34500
	gtagttttgc tcagaatgat ggtttccacc ttcattccatg tccctgcaaa gacatgaact	34560
	catcattttt atggctgcat agtattccat ggtgtatatg tgccacattt tctttatcca	34620
35	ttatatcgct gatggccatt tgggttggtt ccaagtcttt gctattgtga atagtgccac	34680
	aataaacata cgtgtgcacg tgtctttata gtagaatgat ttctaattct ttgggtatat	34740
	accagtaat gggattgctg ggtcaaacag tatttctggt tctagatcct tgaggaatcg	34800
40	ccacactgtc ttccacaatg gttgaactaa tttacacacc catcaacagt gtaaaatttt	34860
	tcctattctt ccacatctc tccagcacct tttgtttcct gactttttta taattgccat	34920
	tctaactggc atgagatggg atctcattgt ggttttgatt tgcatttctc taatgaccag	34980
45	tgatgatgag cttcttttca tgtgtttctt ggccacataa atgacttctt tagagaagca	35040
	tctgttcata tcctttgtcc actttttgat ggggtcgtta ggttttttct tgtaaatgtg	35100
	ttgaagtctt ttgtagattt tggatgtag ccctttgtca gatggataga ttgcaaaaat	35160
50	tttctcccat tctgtagggt gcctgttcac tctgatgata gtcttttgct gtgcagaagc	35220
	tcttttagttt aattagatcc catatgtcaa ttttggcctt tgttgtcatt gcttttgatg	35280
55	ttttagtcgt gaatttttgc ccatgcctat gtcctgaatg gtattgccta gggtatcttc	35340
	taggattttt atgggttttag gttgcacatt taagtcttta atccaccttg agttaatttt	35400

EP 3 620 521 A1

	tgtataaggt gtaaggaagg ggtacagttt cagttttatg catattgcta gccagttttt	35460
	ccagcaccat ttattaaata gggaattctt tctccattgc ttttgtgatg tttgtcaaag	35520
5	atcagatggt cgtagatgtg tggcattatt tctgaggctt ctgttctgtt ccactggtct	35580
	atatatctgt tttggtacca gtaccatgct gtttttgtta ctgtagcctt gtagtatagt	35640
	ttgaagtcag gtagcatcat gcctccagct ttgttctttt tgtttaggat tgtcttggt	35700
10	atatgggctc ttttttgatt ccatatgaca tttaaagtag ttttttctaa ttctttgaaa	35760
	aaagtcagtg gtagcttgat ggggatagca ttgaatctat aaattacttt gggcagtatg	35820
	gccattttta agatattgat tctttctatc tatgagcatg gaatgttttt ccatttgttt	35880
15	gtgtcctctc ttatttcctt gagcagtgag tggtttgtag ctctccttga agaggttctt	35940
	cacatccctt agaagttgta tttctaggta ttttatttta ttctctttgc agcaattgtg	36000
	aatgggagtt caccatgat ttggctctct gcttgtctat tattggtgta taggaacgct	36060
20	tgtgatttct gcacactgat tttgtatctt gagactttgc tgaagctggt tatcagctta	36120
	agattttggg ctgagatgac aggttcttct aaatatacaa tcatgtcatc tgcaaacaga	36180
25	gacaatttga ctctctctct tcctatttga atatgcttta tttctttctc ttgcctgatt	36240
	gtcctggcga gaacttccaa tactatgttg agtaagagtg gcgagagggc atccttgtct	36300
	tgtgccggtt ttcaaagcaa atgattttta aatttccatc ttgatttcat tgttgaccca	36360
30	atgatcattc aggagcaggt tatttaattt ccctgtattt gcatggtttt gaaggttcct	36420
	tttgtagttg atttccaatt ttattctact gtggtctgag agagtgcttg atataatttc	36480
35	aattttttaa aatttattga ggcttgtttt gtggcatatc atatggccta tcttgagaa	36540
	agttccatgt gctgatgaat agaatgtgta ttctgcagtt gttgggtaga atgtcctgta	36600
	aatatctggt aagtccattt gttcttttaa tccattgttt cttttagtagc tgtcttgatg	36660
40	acctgcctag tgcagtcagt ggagtattga agtccccac tattattatg ttgctgtcta	36720
	gtctagtagt aattgtttta taaatttggg atctccagta ttagatgcat atatattaag	36780
	aattgtaata ttctccatt ggacaagggc ttttatcatt atatgatgtc cctctttgtc	36840
45	ttttttaact gctgtttctt taaagtttgt tttgtctgac ataagaatag ctgctttggc	36900
	tcgcttttgg tgtccatttg tgtggaatgt cattttccac ccctttacct taagtttatg	36960
	tgagtcctta tgtgttaggt gagtctcctg aaggcggcag ataactggtt ggtgaattct	37020
50	tattcattct gcaattctgt atcttttaag tggagcattt agtccattta cattcaacat	37080
	cagtattgag gtgtgaggta ctattccatt ctctgtggta tttgttgctt gtgtatcttt	37140
55	ttatctgtat ttttgttgta tatgtcctat gggatttatg ctttaaagag gttctgtttt	37200
	gatgtgcttc cagggtttat ttcaagattt agagctcctt ttatcagttc ttgtagtgtt	37260

EP 3 620 521 A1

	ggcttggttag	tgccgaattc	tctcagcatt	tgtttttctg	aaaaacactg	tgtattttct	37320
	tcatttgtga	agcttagttt	cactggatat	aaaattcttg	gctgataatt	gttttgttta	37380
5	agaaggctga	agatagggcc	atattcactt	ctagctttta	cggtttctgc	tgagaaatct	37440
	gctgttaatc	tgataggttt	tctttcatag	gttacctggt	agtttcacct	cacagctctt	37500
	aagattctct	ttgtcttttag	ataacttttg	atactctgat	gacaatgtac	ctaggcaatg	37560
10	atatttttgc	aatgaatttc	ccaggtgttt	attgagcttc	ttgtatttgg	atatctaggt	37620
	ctctagcaag	gtgggggaag	ttttccttga	ttatttccct	ggataagttt	tccaaacttt	37680
	tagattttctc	ttctttctca	ggaatgctga	ttattcttag	gtttgattgt	ttaacataat	37740
15	cccagatttc	ttggaggctt	tgttcatatt	ttcttattct	tttttctttg	tctttgttgg	37800
	attgggttaa	ttcaaaaact	ttgtcttcaa	gctctgaatt	tcttctgctt	ggattctatt	37860
	gctgagactt	tctagagcat	tttgcatttc	tataagtga	tccattcatc	cattgtttcc	37920
20	tgaagttttg	aatgtttttt	atztatgcta	tctctttaac	tgaagatttc	tcccctcatt	37980
	tcttgtatca	tatttttggg	ttttttaaaa	ttggacttca	ccttcctcgg	atgcctcctt	38040
	gattagctta	ataactgacc	ttctgaatta	tttttcaggt	aaatcagggg	tttcttcttg	38100
25	gtttggatgc	attgctgggtg	agctagtatg	attttttggg	gggtgttaaa	gaaccttggt	38160
	tttcatatta	ccagagttag	ttttctgggt	ccttctcact	tgggtagggt	ctgtcagagg	38220
	gaaagtctag	gcctcaaggc	tgagactttt	gtcccatgag	gtgttccctt	gatgtagcac	38280
30	agtccccctt	ttcctaggcg	tggggcttcc	tgagagccga	actgtagtga	ttgttatctc	38340
	tcttctggat	ctagccaccc	atcaggtcta	ccagactcca	ggctggtact	gggtttgtc	38400
35	tgcacagagt	cttgtgacgt	gaaccatctg	tgggtctctc	agccatagat	acaaccacct	38460
	gtccaatgg	aggtggcaga	ggatgaaatg	gactctgtga	gggtccttac	ttttggttgt	38520
	tcaatgcact	atttttgtgc	tggttggcct	cctgccagga	ggtggcactt	tctagaaagc	38580
40	atcagcagag	gcagtcaggt	ggtggtggct	gggggggctg	gggcacccta	gaactcccaa	38640
	gaatatatgc	cctttgtctt	cagctaccag	ggtgagtaag	gaaggaccat	caggtggggg	38700
	caggactagt	cgtgtctgag	ctcagagtct	ccttgggcag	gtctttctgt	ggctactgtg	38760
45	ggaggatggg	ggtgtagttt	ccaggtcaat	ggatttatgt	tcctaggaca	attatggctg	38820
	cctctgctgt	gtcatgcagg	tcatcaggaa	agtgggggaa	agcaagcagt	cacgtgactt	38880
	gcccagctcc	catgcaactc	aaaaggttgg	tctcacttcc	agcgtgcacc	ctccccgcga	38940
50	acagcaccga	atctgtttcc	atgcagtcag	tgagcaaggc	tgagaacttg	ccccaggcta	39000
	ccagctgcga	aaccaagtag	ggctgtccta	cttccttgcc	agtggagtct	gcacaccaaa	39060
55	ttcatgtccc	cccaccaacc	ccccactgc	ccagccccta	gatctggcca	ggtggagatt	39120
	ttctttttcc	tgtcatcttt	tcccagttcc	tctggcagcc	ctcccaaatg	accctgtga	39180

EP 3 620 521 A1

	ggcaaggcag	aaatggcttc	ctaggggacc	cagagagccc	acagggcttt	tcccgcctgt	39240
	tcctctaccc	ctgtattttg	cttggccctc	taaattgact	cagctccagg	taaggtcaga	39300
5	atctttctct	gtgggtctaga	tcttcaggtt	ccccagtgag	gatgtgtgtt	tgggggtaga	39360
	cgggtccccct	tttccacttc	cacagtttgg	gcactcacia	tatttgggggt	gtttcccggg	39420
	tcctgcagga	gcaatctgct	tctttcagag	gggtgtgtgcg	ttctctcagc	tttcttgatt	39480
10	tattttctgca	gggtggttctg	caaaaaaat	tcctgatggg	agacttcaca	tgctgctctg	39540
	tgcatccgag	tgggagctgc	aatgtacttc	tgctgcctcc	catctgccat	cacctctaa	39600
15	tttgtcggta	atatgcattt	ttaatcaatc	tttttttctc	tctctctctt	tttcttctcc	39660
	cccaaaacta	tactgccctt	tgatatcaag	gaatcaagga	cgtgatgttg	aggggtgggc	39720
	agtggatata	ctctttaccc	cttagggagc	tatatctaga	tttagatatt	gccaatcaaa	39780
20	gataacttaa	ttgaaagcaa	attcataatg	aatacacaca	cacacacaca	catctgcatg	39840
	acaagatfff	taatagttga	aagaataact	aataattgtc	cacaggcaat	aagggtcttt	39900
	taagcaaaac	agttgtgata	aacaggtcat	tcttagaata	gtaatccagc	caatagtaca	39960
25	ggttgcttag	agattatgtc	attaccagag	ttaaaattct	ataatggctt	ctcactccct	40020
	accactgagg	acaagtttat	gtccttaggt	ttatgcttcc	ctgaaacaat	accacctgct	40080
	attctccact	ttacatatca	acggcactgg	ttcttttatct	aactctctgg	cacagcagga	40140
30	gtttgttttc	ttctgcttca	gagctttgaa	tttactatft	cagcttctaa	actttatttg	40200
	gcaatgcctt	cccatggcag	attccttctg	tcattttgcc	tctgttcgaa	tactttctcc	40260
35	ttaatttcat	tcttagttaa	taatattctga	aattattttg	ttgtttaact	taattattaa	40320
	ttttatgtat	gttctacct	gattataatc	ttcagaggaa	agttttattc	tctgacttat	40380
	ttaacttaaa	tgcccactac	tttaaaaatt	atgacattta	tttaacagat	atttgctgaa	40440
40	caaagtgttg	aaaatacatg	ggaaagaatg	cttgaaaaca	cttgaaattg	cttgtgtaaa	40500
	gaaacagttt	tatcagttag	gatttaatac	atgtcagaag	caatgatata	ggaaaaatcg	40560
	aggaataaga	cagttatgga	taaggagaaa	tcaacaaact	cttaaaagat	attgcctcaa	40620
45	aagcataaga	ggaaataagg	gtttatacat	gactttttaga	acactgcctt	ggtttttgga	40680
	taaatgggga	agttgtttga	aaacaggagg	gacccatagat	attccttagt	ctgaggagga	40740
	gcaattaaga	ttcacttggt	tagaggctgg	gagtggtggc	tcacgcctgt	aatcccagaa	40800
50	ttttgggagg	ccaaggcagg	cagatcacct	gaggtcaaga	gttcaagacc	aacctggcca	40860
	acatggtgaa	atcccatctc	tacaaaaata	caaaaattag	acaggcatga	tggcaagtgc	40920
55	ctgtaatccc	agctacttgg	gaggctgagg	aaggagaatt	gcttgaacct	ggaaggcagg	40980
	agttgcagtg	agccgagatc	ataccactgc	actccagcct	gggtgacaga	acaagactct	41040

EP 3 620 521 A1

	gtctcaaaaa	aaaaaaagag	agattcaaaa	gattcacttg	tttaggcctt	agcgggctta	41100
	gacaccagtc	tctgacacat	tcttaaaggt	caggctctac	aaatggaacc	caaccagact	41160
5	ctcagatatg	gccaaagatc	tatacacacc	catctcacag	atcccctatc	ttaaagagac	41220
	cctaatttgg	gttcacctca	gtctctataa	tctgtaccag	cataccaata	aaaatctttc	41280
	tcacccatcc	ttagattgag	agaagtcact	tattattatg	tgagtaactg	gaagatactg	41340
10	ataagttgac	aaatcttttt	ctttcctttc	ttattcaact	tttattttta	cttccaaaga	41400
	acaagtgcaa	tatgtgcagc	tttggtgcgc	aggtcaacat	gtatctttct	ggtcttttag	41460
	ccgcctaaca	ctttgagcag	atataagcct	tacacaggat	tatgaagtct	gaaaggattc	41520
15	caccaatatt	attataattc	ctatcaacct	gatagggttag	gggaaggtag	agctctcctc	41580
	caataagcca	gatttccaga	gtttctgacg	tcataatcta	ccaaggcat	ggatcgagtt	41640
	cagagaaaaa	acaaaagcaa	aaccaaaccct	acaaaaaat	aaaaatccca	aagaaaaaat	41700
20	aaagaaaaaa	acagcatgaa	tacttcctgc	catgttaagt	ggccaatatg	tcagaaacag	41760
	cactgagtta	cagataaaga	tgtctaaact	acagtgcacat	cccagctgtc	acagtgtgtg	41820
25	gactattagt	caataaaaca	gtccctgcct	cttaagagtt	gttttccatg	caaatacatg	41880
	tcttatgtct	tagaataaga	ttccctaaga	agtgaaccta	gcatttatac	aagataatta	41940
	attctaatacc	atagtatctg	gtaaagagca	ttctaccatc	atctttaccg	agcatagaag	42000
30	agctacacca	aaaccctggg	tcatcagcca	gcacatacac	ttatccagtg	ataaatacac	42060
	atcatcgggt	gcctacatac	atacctgaat	ataaaaaaaaa	tacttttgct	gagatgaaac	42120
	aggcgtgatt	tatttcaa	aggtacggat	aagtagatat	tgaagtaagg	attcagtcct	42180
35	atattatatt	acataacatt	aatctattcc	tgcactgaaa	ctgttgcttt	ataggatttt	42240
	tcactacact	aatgagaact	taagagataa	tggcctaaaa	ccacagagag	tatattcaaa	42300
	gataagtata	gcacttctta	tttggaacc	aatgcttact	aaatgagact	aagacgtgtc	42360
40	ccatcaaaaa	tcctggacct	atgcctaaaa	cacatttcac	aatccctgaa	cttttcaaaa	42420
	attggtacat	gctttaactt	taaactacag	gcctcactgg	agctacagac	aagaagggtga	42480
45	aaaacggctg	acaaaagaag	tcctggatc	ttctatgggtg	ggagaagaaa	actagctaaa	42540
	gggaagaata	aattagagaa	aaattggaat	gactgaatcg	gaacaaggca	aaggctataa	42600
	aaaaaattaa	gcagcagtat	cctcttgggg	gccccctccc	cacactatct	caatgcaaat	42660
50	atctgtctga	aacgggtccct	ggctaaactc	cacccatggg	ttggccagcc	ttgccttgac	42720
	caatagcctt	gacaaggcaa	acttgaccaa	tagtcttaga	gtatccagtg	aggccagggg	42780
	ccggcggctg	gctagggatg	aagaataaaa	ggaagcacc	ttcagcagtt	ccacacactc	42840
55	gcttctggaa	cgtctgaggt	tatcaataag	ctcctagtcc	agacgccatg	ggtcatttca	42900
	cagaggagga	caaggctact	atcacaagcc	tgtggggcaa	ggtgaatgtg	gaagatgctg	42960

EP 3 620 521 A1

	gaggagaaac cctgggaagg taggctctgg tgaccaggac aaggaggaggga aggaaggacc	43020
	ctgtgcctgg caaaagtcca ggctcgcttct caggatttgt ggcaccttct gactgtcaaa	43080
5	ctgttcttgt caatctcaca ggctcctggg tgtctaccca tggaccacaga ggttctttga	43140
	cagctttggc aacctgtcct ctgcctctgc catcatgggc aaccccaaag tcaaggcaca	43200
	tggcaagaag gtgctgactt ccttgggaga tgccataaag cacctggatg atctcaaggg	43260
10	cacctttgcc cagctgagtg aactgcactg tgacaagctg catgtggatc ctgagaactt	43320
	caagggtgagt ccaggagatg tttcagcact gttgccttta gtctcgaggc aacttagaca	43380
15	actgagtatt gatctgagca cagcagggtg tgagctgttt gaagatactg gggttgggag	43440
	tgaagaaact gcagaggact aactgggctg agaccagtg gcaatgtttt agggcctaag	43500
	gagtgcctct gaaaatctag atggacaact ttgactttga gaaaagagag gtggaaatga	43560
20	ggaaaatgac ttttctttat tagatttcgg tagaaagaac tttcaccttt cccctatttt	43620
	tgttattcgt tttaaaacat ctatctggag gcaggacaag tatggtcatt aaaaagatgc	43680
	aggcagaagg catatattgg ctcaagtcaa gtgggggaact ttggtggcca aacatacatt	43740
25	gctaaggcta ttcctatatc agctggacac atataaaatg ctgctaattgc ttcattacaa	43800
	acttatatcc ttttaattcca gatgggggca aagtatgtcc aggggtgagg aacaattgaa	43860
	acatttgggc tggagtagat tttgaaagtc agctctgtgt gtgtgtgtgt gtgtgtgcgc	43920
30	gcgtgtgttt gtgtgtgtgt gagagcgtgt gtttctttta acgttttcag cctacagcat	43980
	acagggttca tgggtggcaag aagataacaa gatttaaatt atggccagtg actagtgtctg	44040
35	caagaagaac aactacctgc atttaatggg aaagcaaat ctcaaggctt gagggaggtt	44100
	aacataggct tgattctggg tggaagcttg gtgtgtagtt atctggaggc caggctggag	44160
	ctctcagctc actatgggtt catctttatt gtctcctttc atctcaacag ctctgggaa	44220
40	atgtgctggt gaccgttttg gcaatccatt tcggcaaaga attcaccctt gaggtgcagg	44280
	cttcctggca gaagatggtg actggagtgg ccagtgcctt gtctccaga taccactgag	44340
	ctcactgccc atgatgcaga gctttcaagg ataggcttta ttctgcaagc aatcaaataa	44400
45	taaatctatt ctgctaagag atcacacatg gttgtcttca gttctttttt tatgtctttt	44460
	taaatatatg agccacaaag ggttttatgt tgagggatgt gtttatgtgt atttatacat	44520
	ggctatgtgt gtttgtgtca tgtgcacact ccacactttt ttgtttacgt tagatgtggg	44580
50	ttttgatgag caaataaaag aactaggcaa taaagaaact tgtacatggg agttctgcaa	44640
	gtgggagtaa aagggtgcagg agaaatctgg ttggaagaaa gacctctata ggacaggact	44700
55	cctcagaaac agatgttttg gaagagatgg ggaaagggtt agtgaagggg gctgaacccc	44760
	cttccttggg ttgcagcaca gcagcgagga aggggctcaa cgaagaaaaa gtgttccaag	44820

EP 3 620 521 A1

	cttttaggaag tcaaggttta ggcaggata gccattctat tttatttaggg gcaatactat	44880
	ttccaacggc atctggcttt tctcagccct tgtgaggctc tacagggagg ttgaggtggt	44940
5	agagatcaga gcaggaaaca gggttttctt tccacggtaa ctacaatgaa gtgatcctta	45000
	ctttactaag gaacttttca ttttaagtgt tgacgcatgc cttaaagagg gaaattaatc	45060
	ccataccctt aagtctacag actggtcaca gcatttcaag gaggagacct cattgtaagc	45120
10	ttctagggag gtggggactt aggtgaagga aatgagccag cagaagctca caagtcagca	45180
	tcagcgtgtc atgtctcagc agcagaacag cacggtcaga tgaaaatata gtgtgaagaa	45240
	tttgtataac attaatgag aaggcagatt cactggagtt cttatataat tgaaagttaa	45300
15	tgcacgttaa taagcaagag tttagtttaa tgtgatggtg ttatgaactt aacgcttggtg	45360
	tctccagaaa attcacatgc tgaatcccca actccaatt ggctccattt gtgggggagg	45420
	ctttggaaaa gtaatcaggt ttagaggagc tcatgagagc agatcccat catagaatta	45480
20	ttttcctcat cagaagcaga gagattagcc atttctcttc cttctggtga ggacacagtg	45540
	ggaagtcagc cacctgcaac ccaggaagag agccctgacc aggaaccagc agaaaagtga	45600
25	gaaaaaatcc tgttgttgaa gtcaccagc ctatgctatt ttgttatagc accttgcaact	45660
	aagtaaggca gatgaagaaa gagaaaaaaa taagcttcgg tgttcagtggtg attagaaacc	45720
	atgtttatct cagggtttaca aatctccact tgtcctctgt gtttcagaat aaaataccaa	45780
30	ctctactact ctcatctgta agatgcaaata agtaagcctg agcccttctg tctaactttg	45840
	aattctatct tttcttcaac gtactttagg cttgtaatgt gtttatatac agtgaaatgt	45900
	caagttcttt ctttatatct ctttcttct ttttttctc cagcctcaga gttttccaca	45960
35	tgcccttctc actttcagga acttcttct ccaaactgtc tctgcctggc tccatcaaata	46020
	cataaaggac ccacttcaaa tgccatcact cactaccatt tcacaattcg cactttcttt	46080
	ctttgtcctt ttttttttta gtaaaacaag tttataaaaa attgaaggaa taaatgaatg	46140
40	gctacttcat aggcagagta gacgcaaggg ctactgggtg ccgattttta ttgttatctt	46200
	tcaatagtat gctaaacaag gggtagatta tttatgctgc ccatttttag accataaaag	46260
	ataacttctc gatgttgcca tggcattttt ttctttttaa ttttatttca tttcatttta	46320
45	atttcgaagg tacatgtgca ggatgtgcag gcttggttaca tgggtaaatg tgtgtctttc	46380
	tggcctttta gccatctgta tcaatgagca gatataagct ttacacagga tcatgaagga	46440
	tgaaagaatt tcaccaatat tataataatt tcaatcaacc tgatagctta ggggataaac	46500
50	taatttgaag atacagcttg cctccgataa gccagaattc cagagcttct ggcattataa	46560
	tctagcaagg ttagagatca tggatcactt tcagagaaaa aaaaaaaca actaaccaaa	46620
55	agcaaaacag aacaaaaaaa ccaccataaa tacttctac cctgttaatg gtccaatatg	46680
	tcagaaacag cactgtgtta gaaataaagc tgtctaaagt aactaatat tcgagttata	46740

EP 3 620 521 A1

	atagtgtgtg	gactattagt	caataaaaac	aacccttgcc	tcttttagagt	tgttttccat	46800
	gtacacgcac	atcttatgtc	ttagagtaag	attccctgag	aagtgaacct	agcatttata	46860
5	caagataatt	aattctaatac	cacagtacct	gccaaagaac	attctaccat	catctttact	46920
	gagcatagaa	gagctacgcc	aaaaccctgg	gtcatcagcc	agcacacaca	cttatccagt	46980
	ggtaaatata	catcatctgg	tgtatacata	catacctgaa	tatgggaatca	aatatttttc	47040
10	taagatgaaa	cagtcgatgat	ttattttcaaa	taggtacgga	taagtagata	ttgaggtaag	47100
	cattaggtct	tatattatgt	aacactaatac	tattactgog	ctgaaactgt	ggcttttatag	47160
	aaattgtttt	cactgcacta	ttgagaaatt	aagagataat	ggcaaaagtc	acaaagagta	47220
15	tattcaaaaa	gaagtatagc	actttttcct	tagaaaccac	tgctaactga	aagagactaa	47280
	gatttgtccc	gtcaaaaatc	ctggacctat	gcctaaaaca	catttcacaa	tcctgaact	47340
20	tttcaaaaat	tggtacatgc	tttagcttta	aactacaggc	ctcactggag	ctagagacaa	47400
	gaaggtaaaa	aacggctgac	aaaagaagtc	ctggtatcct	ctatgatggg	agaaggaaac	47460
	tagctaaagg	gaagaataaa	ttagagaaaa	actggaatga	ctgaatcgga	acaaggcaaa	47520
25	ggctataaaa	aaaattagca	gtatcctctt	gggggcccct	tccccacact	atctcaatgc	47580
	aaatatctgt	ctgaaacggg	ccctggctaa	actccaccca	tgggttggcc	agccttgcc	47640
	tgaccaatag	ccttgacaag	gcaaacttga	ccaatagtct	tagagtatcc	agtgaggcca	47700
30	ggggccggcg	gctggctagg	gatgaagaat	aaaaggaagc	acccttcagc	agttccacac	47760
	actcgcttct	ggaacgtctg	aggttatcaa	taagctccta	gtccagacgc	catgggtcat	47820
35	ttcacagagg	aggacaaggc	tactatcaca	agcctgtggg	gcaaggtgaa	tgtggaagat	47880
	gctggaggag	aaaccctggg	aaggtaggct	ctggtgacca	ggacaaggga	gggaagggaag	47940
	gaccctgtgc	ctggcaaaaag	tccaggctgc	ttctcaggat	ttgtggcacc	ttctgactgt	48000
40	caaactgttc	ttgtcaatct	cacaggctcc	tggttgtcta	cccatggacc	cagaggttct	48060
	ttgacagctt	tggcaacctg	tcctctgcct	ctgccatcat	gggcaacccc	aaagtcaagg	48120
	cacatggcaa	gaagggtgctg	acttccttgg	gagatgccac	aaagcacctg	gatgatctca	48180
45	agggcacctt	tgcccagctg	agtgaactgc	actgtgacaa	gctgcatgtg	gatcctgaga	48240
	acttcaaggt	gagtccagga	gatgtttcag	ccctgttgcc	tttagtctcg	aggcaactta	48300
	gacaacggag	tattgatctg	agcacagcag	gggtgtgagct	gtttgaagat	actgggggttg	48360
50	gggggtgaaga	aactgcagag	gactaactgg	gctgagaccc	agtggtaatg	ttttagggcc	48420
	taaggagtgc	ctctaaaaat	ctagatggac	aattttgact	ttgagaaaag	agagggtggaa	48480
55	atgaggaaaa	tgacttttct	ttattagatt	ccagtagaaa	gaactttcat	ctttccctca	48540
	tttttgttgt	tttaaaacat	ctatctggag	gcaggacaag	tatgggtcgtt	aaaaagatgc	48600

EP 3 620 521 A1

	aggcagaagg	catatatattgg	ctcagtc meta	gtgggggaact	ttggtggcca	aacatacatt	48660
	gctaaggcta	ttcctatatc	agctggacac	atataaaatg	ctgctaatac	ttcattacaa	48720
5	acttatatcc	tttaattcca	gatgggggca	aagtatgtcc	aggggtgagg	aacaattgaa	48780
	acattttggc	tggagtagat	tttgaaagtc	agctctgtgt	gtgtgtgtgt	gtgtgcgcgc	48840
	gcgcgtgtgt	gtgtgtgtgt	cagcgtgtgt	ttcttttaac	gtcttcagcc	tacaacatac	48900
10	agggttcatg	gtggcaagaa	gatagcaaga	tttaaattat	ggccagtgac	tagtgcttga	48960
	aggggaacaa	ctacctgcat	ttaatgggaa	ggcaaaatct	caggctttga	gggaagttaa	49020
	cataggcttg	attctgggtg	gaagcttggt	gtgtagttat	ctggaggcca	ggctggagct	49080
15	ctcagctcac	tatgggttca	tctttattgt	ctcctttcat	ctcaacagct	cctgggaaat	49140
	gtgctggtga	ccgttttggc	aatccatttc	ggcaaagaat	tcaccctga	ggtgcaggct	49200
	tcctggcaga	agatggtgac	tgagtggtgc	agtgcctgt	cctccagata	ccactgagct	49260
20	cactgcccac	gattcagagc	tttcaaggat	aggctttatt	ctgcaagcaa	tacaaataat	49320
	aaatctattc	tgctgagaga	tcacacatga	ttttcttcag	ctcttttttt	tacatctttt	49380
25	taaatatatg	agccacaaag	ggtttatatt	gagggaagtg	tgtatgtgta	tttctgcatg	49440
	cctgtttgtg	tttgtggtgt	gtgcatgctc	ctcattttatt	tttatatgag	atgtgcattt	49500
	tgatgagcaa	ataaaagcag	taaagacact	tgtacacggg	agttctgcaa	gtgggagtaa	49560
30	atggtgtagg	agaaatccgg	tgggaagaaa	gacctctata	ggacaggact	tctcagaaac	49620
	agatgttttg	gaagagatgg	gaaaaggttc	agtgaagacc	tgggggctgg	attgattgca	49680
	gctgagtagc	aaggatgggt	cttaagggaag	ggaaagtgtt	ccaagcttta	ggaattcaag	49740
35	gtttagtcag	gtgtagcaat	tctattttat	taggaggaat	actattttcta	atggcactta	49800
	gcttttcaca	gcccttgtgg	atgcctaaga	aagtgaat	aatcccatgc	cctcaagtgt	49860
	gcagattggt	cacagcattt	caaggagag	acctcattgt	aagactctgg	gggaggtggg	49920
40	gacttaggtg	taagaaatga	atcagcagag	gtccacaagt	cagcatgagc	atgttatgtc	49980
	tgagaaacag	accagcactg	tgagatcaaa	atgtagtggg	aagaatttgt	acaacattaa	50040
	ttggaaggct	tacttaatgg	aatttttgtg	tagttggatg	ttagtgcatc	tctataagta	50100
45	agagtttaat	atgatggtgt	tacggaccta	atgtttgtgt	ctcctcaaaa	ttcacatgct	50160
	gaatcccca	ctcccaactg	accttatctg	tgggggaggc	ttttgaaaag	taattaggtt	50220
	tagatgagct	cataagagca	gatccccatc	ataaaattat	tttccttata	agaagcagag	50280
50	agacaagcca	tttctctttc	ctcccgttga	ggacacagtg	agaagtccgc	catctgcaat	50340
	ccaggaagag	aaccctgacc	acgagtcagc	cttcagaaat	gtgagaaaaa	actctgttgt	50400
55	tgaagccacc	cagtcttttg	tattttgtta	tagcaccttg	cactgagtaa	ggcagatgaa	50460
	gaaggagaaa	aaaataagct	tgggttttga	gtggactaca	gaccatgttt	atctcagggt	50520

EP 3 620 521 A1

	tgcaaagctc ccctcgtccc ctatgtttca gtataaaata cctactctac tactctcatc	50580
	tataagaccc aaataataag cctgcgcccct tctctctaac tttgatttct cctattttta	50640
5	cttcaacatg ctttactcta gccttgtaat gtctttacat acagtgaaat gtaaagttct	50700
	ttattctttt tttctttctt tcttttttct cctcagcctc agaatttgge acatgccctt	50760
	ccttctttca ggaacttctc caacatctct gcctggctcc atcatatcat aaagggtcca	50820
10	cttcaaagtc agtcactacc gtttcagaat atgcactttc tttctttttt gttttttgtt	50880
	ttttttaagt caaagcaaag ttcttgagag agtaaagaaa taaacgaatg actactgcat	50940
	aggcagagca gccccgaggg ccgctgggtg ttccctttat gggtatttct tgatgatatg	51000
15	ttaaacaagt ttgggattat ttatgccttc tctttttagg ccatataggg taactttctg	51060
	acattgccat ggcatTTTTT ttttaattta atttactgtt accttaaatt caggggtaca	51120
20	cgtacaggat atgcagggtt gttttatagg taaaagtgtg ccatgggttt aatgggtttt	51180
	ttttttcttg taaagtgtgt taagtttctt gtttactctg gatattaggg ctttgtcaga	51240
	agaatagatt ggaaaatctt tttccatttc tgtagattgt ctttcgctct gatggtagtt	51300
25	tcttttgctg agcaggagct ctttagttta attagattcc attgggtcaat ttttgctttt	51360
	gctgcaattg cttttcacgc tttcatcatg aaatctgtgc ccgtgtttat atcatgaata	51420
	gtattgcctt gatttttttc taggcttttt atagtttggg gtttttcatt taagtctcta	51480
30	atccatctgg agttaatttt ggataaggta taaggaagga gtccagtttc atttttcagc	51540
	atatggctag ccagttctcc cccatcattt attaaattga aaatcctttc cccattgctt	51600
35	gcttttgtca ggtttctaaa agaccagatg gttgtaggta caatatgcag tttcttcaag	51660
	tcatataata ccatctgaaa tctcttatta attcatttct tttagtatgt atgctggtct	51720
	cctctgctca ctatagttag ggcaccatta gccagagaat ctgtctgtct agttcatgta	51780
40	agattctcag aattaagaaa aatggatggc atatgaatga aacttcatgg atgacatatg	51840
	gaatctaata tgtatttggt gaattaatgc ataagatgca acagagagaa gttgacaact	51900
	gcaatgataa cctgggtattg atgatataag agtctataga tcacagtaga agcaataatc	51960
45	atggaaaaca attggaaatg gggaacagcc acaacaaga aagaatcaat acttcagga	52020
	aagtgactgc aggtcacttt tcctggagcg ggtgagagaa aagtggaagt tagcagtaac	52080
	tgctgaattc ctgggtgggt gatggaaaga tggggcagct gttcactggg acgcagggtt	52140
50	ttagatgtat gtacctaaag atatgaggta tggcaatgaa cagaaattct tttgggaatg	52200
	agttttaggg ccattaaagg acatgacctg aagtttctc tgaggccagt cccacaact	52260
55	caatataaat gtgtttcctg catatagtca aagttgccac tttttttct tcatatcatc	52320
	gatctctgct cttaaagata atcttggttt tgccctcaaac tgtttgtcac tacaaacttt	52380

EP 3 620 521 A1

	ccccatgttc	ctaagtaaaa	caggtaactg	cctctcaact	atatcaagta	gactaaaata	52440
	tttgtgtctct	aatatcagaa	attcagcttt	aatatattgg	gtttaactct	ttgaaattta	52500
5	gagtctcctt	gaaatacaca	tgggggtgat	ttcctaaact	ttatttcttg	taaggattta	52560
	tctcaggggt	aacacacaaa	ccagcatcct	gaacctctaa	gtatgaggac	agtaagcctt	52620
	aagaatataa	aataaactgt	tcttctctct	gccggtggaa	gtgtgccctg	tctattcctg	52680
10	aaattgcttg	tttgagacgc	atgagacgtg	cagcacatga	gacacgtgca	gcagcctgtg	52740
	gaatattgtc	agtgaagaat	gtctttgcct	gattagatat	aaagacaagt	taaacacagc	52800
	attagactat	agatcaagcc	tgtgccagac	acaaatgacc	taatgccag	cacgggccac	52860
15	ggaatctcct	atcctcttgc	ttgaacagag	cagcacactt	ctcccccaac	actattagat	52920
	gttctggcat	aattttgtag	atatgtagga	tttgacatgg	actattgttc	aatgattcag	52980
	aggaaatctc	ctttgttcag	ataagtacac	tgactactaa	atggattaaa	aaacacagta	53040
20	ataaaaacca	gttttccct	tacttcccta	gtttgtttct	tattctgctt	tcttccaagt	53100
	tgatgctgga	tagaggtgtt	tatttctatt	ctaaaaagtg	atgaaattgg	ccgggcgcgg	53160
25	tggctcacac	ctgtaatccc	agcactttgg	gaggctgagg	tgggcggatc	acgaggtcag	53220
	gagatcaaga	ccatcctggc	taacatggtg	aaaccccatc	tctactaaaa	atacaaaaaa	53280
	ttagccagag	acagtggcgg	gtgcctgtag	tcccagctac	tccggaggct	gaggcaggag	53340
30	aatggcgtga	acctgggagg	cagagcttgc	ggtgagcaga	gatcgcgcca	ctgcacactc	53400
	cagcctgggt	gacaaagcga	gactccatct	caaaaaaaaa	aaaaaaaaaa	agaaaaagaa	53460
	agaaagaaag	aaaaaaaaaac	tgatgaaatt	gtgtattcaa	tgtagtctca	agagaattga	53520
35	aaaccaagaa	aggctgtggc	ttcttccaca	taaagcctgg	atgaataaca	ggataaacag	53580
	ttgttacatt	gtcacaactc	ctgatccagg	aattgatggc	taagatattc	gtaattctta	53640
	tccttttcag	ttgtaactta	ttcctatttg	tcagcattca	ggttattagc	ggctgctggc	53700
40	gaagtccctg	agaaataaac	tgcacactgg	atggtggggg	tagtgtagga	aaatggaggg	53760
	gaagggaagta	aagtttcaaa	ttaagcctga	acagcaaagt	tcccctgaga	aggccacctg	53820
	gattctatca	gaaactcgaa	tgtccatctt	gcaaaacttc	cttgcccaaa	ccccaccctt	53880
45	ggagtccaaa	cccacccttg	accaatagat	tcattttact	gagggaggca	aagggtggt	53940
	caatagattc	atttcactgg	gagaggcaaa	gggctggggg	ccagagagga	gaagtaaaaa	54000
	gccacacatg	aagcagcaat	gcaggcatgc	ttctggctca	tctgtgatca	ccaggaaact	54060
50	cccagatctg	acactgtagt	gcatttcact	gctgacaaga	aggctgctgc	caccagcctg	54120
	tgaagcaagg	ttaaggtgag	aaggctggag	gtgagattct	gggcaggtag	gtactggaag	54180
55	ccgggacaag	gtgcagaaag	gcagaaagtg	tttctgaaag	agggattagc	ccgttgtctt	54240
	acatagtctg	actttgcacc	tgctctgtga	ttatgactat	cccacagtct	cctggttgtc	54300

EP 3 620 521 A1

	tacccatgga cctagaggta ctttgaaagt tttggatatc tgggctctga ctgtgcaata	54360
	atgggcaacc ccaaagtcaa ggcacatggc aagaagggtgc tgatctcctt cggaaaagct	54420
5	gttatgctca cggatgacct caaaggcacc tttgctacac tgagtgcct gcaactgtaac	54480
	aagctgcacg tggaccctga gaacttcctg gtgagtagta agtacactca cgctttcttc	54540
	tttaccctta gatatttgca ctatgggtac ttttgaaagc agagggtggct ttctcttggtg	54600
10	ttatgagtca gctatgggat atgatatttc agcagtggga ttttgagagt tatgttgctg	54660
	taaataacat aactaaaatt tggtagagca aggactatga ataatggaag gccacttacc	54720
	atttgatagc tctgaaaaac acatcttata aaaaattctg gccaaaatca aactgagtgt	54780
15	ttttggatga gggaacagaa gttgagatag agaaaataac atctttcctt tggtcagcga	54840
	aattttctat aaaaattaat agtcactttt ctgcatagtc ctggagggtta gaaaaagatc	54900
20	aactgaacaa agtagtggga agctgttaaa aagaggattg tttccctccg aatgatgatg	54960
	gtatactttt gtacgcatgg tacaggattc tttgttatga gtgtttggga aaattgtatg	55020
	tatgtatgta tgtatgtatg tgatgactgg ggacttatcc tatccattac tgttccttga	55080
25	agtactatta tcctactttt taaaaggacg aagtctctaa aaaaaaatg aaacaatcac	55140
	aatatgttgg ggtagtgagt tggcatagca agtaagagaa ggataggaca caatgggagg	55200
	tgcagggctg ccagtcatat tgaagctgat atctagccca taatggtgag agttgctcaa	55260
30	actctggtga aaaaggatgt aagtgttata tctatttact gcaagtccag cttgaggcct	55320
	tctattcact atgtaccatt ttctttttta tcttcaactcc ctcccagct cttaggcaac	55380
35	gtgatattga ttgttttggc aaccacttc agcgaggatt ttaccctaca gatacaggct	55440
	tcttggcagt aactaacaaa tgctgtggtt aatgctgtag ccacaagac cactgagttc	55500
	cctgtccact atgtttgtac ctatggtcca ctatgtttgt acctatgtcc caaaatctca	55560
40	tctccttttag atgggggagg ttggggagaa gagcagtatc ctgcctgctg attcagttcc	55620
	tgcattgataa aaatagaata aagaaatatg ctctctaaga aatatcattg tactcttttt	55680
	ctgtctttat attttaccct gattcagcca aaaggacgca ctatttctga tggaaatgag	55740
45	aatgttggag aatgggagtt taaggacaga gaagatactt tcttgcaatc ctgcaagaaa	55800
	agagagaact cgtgggtgga tttagtgggg tagttactcc taggaagggg aaatcgtctc	55860
	tagaataaga caatgttttt acagaaaggg aggtcaatgg aggtactctt tggagggtga	55920
50	agaggattgt tggtagtgtg tagaggatatg ttaggactca aattagaagt tctgtatagg	55980
	ctattatttg tatgaaactc aggatatagc tcatttggtg actgcagttc acttctactt	56040
55	attttaaaca acatatTTTT tattatttat aatgaagtgg ggatggggct tcctagagac	56100
	caatcaaggg ccaaaccttg aactttctct taacgtcttc aatggtatta atagagaatt	56160

EP 3 620 521 A1

	atctctaagg catgtgaact ggctgtcttg gttttcatct gtacttcac	tgctacctct	56220
	gtgacctgaa acatatttat aattccatta agctgtgcat atgatagatt tatcatatgt		56280
5	attttcctta aaggattttt gtaagaacta attgaattga tacctgtaaa gtctttatca		56340
	cactacccaa taaataataa atctctttgt tcagctctct gtttctataa atatgtacaa		56400
	gttttattgt ttttagtggt agtgatttta ttctctttct atatatatac acacacatgt		56460
10	gtgcattcat aaatatatac aatttttatg aataaaaaat tattagcaat caatattgaa		56520
	aaccactgat ttttgtttat gtgagcaaac agcagattaa aaggctgaga tttaggaaac		56580
	agcacgttaa gtcaagttga tagaggagaa tatggacatt taaaagaggc aggatgatat		56640
15	aaaattaggg aaactggatg cagagaccag atgaagtaag aaaaatagct atcgttttga		56700
	gcaaaaatca ctgaagtttc ttgcatatga gagtgacata ataaataggg aaacgtagaa		56760
	aattgattca catgtatata tatatataga actgattaga caaagtctaa cttgggtata		56820
20	gtcagaggag cttgctgtaa ttatattgag gtgatggata aagaactgaa gttgatggaa		56880
	acaatgaagt taagaaaaa aatcgagtaa gagaccattg tggcagtgat tgcacagaac		56940
25	tggaaaacat tgtgaaacag agagtcagag atgacagcta aaatccctgt ctgtgaatga		57000
	aaagaaggaa atttattgac agaacagcaa atgcctacaa gccccctgtt tggatctggc		57060
	aatgaacgta gccattctgt ggcaatcact tcaaaactct gtaccaaga cccttaggaa		57120
30	gtatgtagca ccctcaaacc taaaacctca aagaaagagg ttttagaaga tataataccc		57180
	tttcttctcc agtttcatta atcccaaac ctctttctca aagtatttcc tctatgtgtc		57240
	caccccaaag agctcacctc accatatctc ttgagtggga gcacatagat aggcggtgct		57300
35	accatctaac agcttctgaa attcctttgt catatttttg agtccccact aataaccac		57360
	aaagcagaat aaataccagt tgctcatgta caataatcac tcaactgctg tctttagca		57420
	tacattaatt aagcacattc tttgaataat tactgtgtcc aaacaatcac actttaaaat		57480
40	ctcacacttg tgctatccct tgcccttctg aatgtcactc tgtattttaa atgaagagat		57540
	gagggttgaa tttcctgtgt tacttattgt tcatctctcg atgaggagt ttcacattca		57600
	ccttttagtgg aaaacacata agtacacac ttacaggaaa aatataccaa actgacatgt		57660
45	agcatgaatg cttgtgcatg tagtcatata aaatcttgta gcaatgtaaa cattctctga		57720
	tatacacata cagatgtgtc tatatgtcta cacaatttct tatgctccat gaacaaacat		57780
	tccatgcaca cataagaaca cacactgtta cagatgcata cttgagtga ttgacaaaat		57840
50	taccccgatc aatctagaga atttgattt ctgcatttga ctctgttagc tttgtacatg		57900
	ctgttcattt actctgggtg atgtctttcc ctcatcttgc cttgtctatc ttgtactcat		57960
55	actttaagtc ctaacttata tgttatctca actaagaagc taaaaaaaa taattttaac		58020
	tgggcttaaa gccctgtcta taaactctgc tacaattatg ggctctttct tataatattt		58080

EP 3 620 521 A1

	agtggtttttc	ctactaatgt	acttaatctg	ctcattgtat	attcctacca	ctaaatttta	58140
	acctcttttta	tggttagagac	attgtcttgt	aaactcttat	ttccctagta	tttggagatg	58200
5	aaaaaaaaga	ttaaattatc	caaaattaga	tctctctttt	ctacattatg	agtattacac	58260
	tatccataga	gaagtttgtt	tgagacctaa	actgaggaac	ctttggttct	aaaatgacta	58320
	tgtgatatct	tagtatattat	aggtcatgag	gttccttctt	ctgcctctgc	tatagtttga	58380
10	ttagtcaaca	agcatgtgtc	atgcatttat	tcacatcaga	atttcataca	ctaataagac	58440
	atagtatcag	aagtcagttt	attagttata	tcagttaggg	tccatcaagg	aaaggacaaa	58500
	ccattatcag	ttactcaacc	tagaattaaa	tacagctctt	aatagttaat	tatccttgta	58560
15	ttggaagagc	taaaatatca	aataaaggac	agtgacagaa	tctagatggt	agtaacatca	58620
	gaaaacctct	tccgccatta	ggcctagaag	ggcagaagga	gaaaatgttt	ataccaccag	58680
20	agtccagaac	cagagcccat	aaccagaggt	ccactggatt	cagtgcagct	gtgggtgctc	58740
	cttggagaga	gccagaactg	tctaattggg	gcatcaaagt	atcagccata	aaaaaccata	58800
	aaaaagactg	tctgctgtag	gagatccgtt	cagagagaga	gagagaccag	aaataatctt	58860
25	gcttatgctt	tccctcagcc	agtgtttacc	attgcagaat	gtacatgcga	ctgaaaggg	58920
	gaggaaacct	gggaaatgtc	agttcctcaa	atacagagaa	cactgagggg	aggatgagaa	58980
	ataaatgtga	aagcagacat	gaatggtaat	tgacagaagg	aaactaggat	gtgtccagta	59040
30	aatgaataat	tacagtgtgc	agtgattatt	gcaatgatta	atgtattgat	aagataatat	59100
	gaaaacacag	aattcaaaca	gcagtgaact	gagattagaa	ttgtggagag	cactggcatt	59160
35	taagaatgtc	acacttagaa	tgtgtctcta	ggcattgttc	tgtgcatata	tcatctcaat	59220
	attcattatc	tgaaaattat	gaattaggta	caaagctcaa	ataatttatt	ttttcaggtt	59280
	agcaagaact	tttttttttt	ttttctgaga	tagagcattg	ctatggttgc	ccaggctgga	59340
40	gtgcaatggc	atgatccagg	ctcactgcaa	catctgcctc	ccaggttcaa	gcgattctcc	59400
	tgcctcagcc	tcccaagtag	ctggcactac	aggcatgtgc	caccaccatg	cctggcta	59460
	tttctatttt	tagtagatag	ggggtttcac	catgttggtc	aggctgatct	cgaactccta	59520
45	acatcaggtg	atccaccctc	ctcggcctct	gaaagtgtct	ggatcacagg	cgtgagccac	59580
	cacaccacgc	caagaatgtg	aattttgtag	aaggatataa	cccatatttc	tctgacccta	59640
	gagtccttag	tatacctccc	ataccatgtg	gtcctatctc	cttacatata	tttcccatct	59700
50	ttcacccctac	cttttctctt	ttgtttcagc	ttttcactgt	gtcaaaatct	agaaccttat	59760
	ctcctacctg	ctctgaaacc	aacagcaagt	tgacttccat	tctaaccac	attggcatta	59820
55	cactaattaa	aatcgatact	gagttctaaa	atcatcgggg	attttgggga	ctatgtctta	59880
	cttcatactt	ccttgagatt	tcacattaaa	tggttggtgtt	cattaaaggt	ccttcattta	59940

EP 3 620 521 A1

	actttgtatt	catcacactc	ttggattcac	agttatatct	aaactcttaa	atacagcctg	60000
	tataatccca	attcccaact	ctgatttcta	acctctgacc	tccaacctca	gtgccaaacc	60060
5	catatatcaa	acaatgtact	gggcttattt	atatagatgt	cctataggca	cctcagactc	60120
	agcatgggta	tttcacttgt	tatactaaaa	ctgtttctct	tccagtgttt	tccatttttag	60180
	tcattagata	gctacttgcc	cattcaccaa	ggtcacagat	taaaatcatt	tccttacctc	60240
10	taatcaacag	ttcgattctg	cttcaatttg	tccctatcta	ttaatcacca	ctcttactgc	60300
	ccagtcaggt	cctcattggt	tcctgaacaa	gagtagatgc	tattctttcc	acttttagac	60360
	cttatcctgg	ctggatgagg	tggctcaggc	ttgtaaaccc	agcactttgg	gaggccaagg	60420
15	caggcagatc	acttgagggt	aggagttcaa	gaccagcctg	accaacatgg	tgaaccccca	60480
	tctctactaa	aaatacaaaa	tcagccgggc	gtgtggtgca	tgctgcagt	cccagctatt	60540
	cagggtggctg	aggcaggaga	attgcttgaa	cccaggaggc	agagggtgcg	gtgagcctag	60600
20	attgcaccat	tgcactctag	cttgggcaat	agggatgaaa	ctccatctca	gaagagaaaa	60660
	gaaaaaaaga	ccttattctg	ttatacaaat	cctctcaatg	caatccatat	agaataaaca	60720
25	tgtaaccaga	tctcccaatg	tgtaaaatca	tttcaggtag	aacagaatta	aagtgaaaag	60780
	ccaagtcttt	ggaattaaca	gacaaagatc	aaataacagt	cctcatggcc	ttaagaattt	60840
	acctaacatt	tttttttagaa	tcaattttct	tatatatgaa	ttggaaacat	aattcctccc	60900
30	tcacaaacac	attctaagat	tttaaggaga	tattgatgaa	gtacatcatc	tgtcattttt	60960
	aacaggtagt	ggtagtgatt	cacacagcac	attatgatct	gttcttgat	gttctgttcc	61020
	attctgtatt	cttgacctgg	ttgtattctt	tctgagctcc	agatccacat	atctaagtac	61080
35	atcttttttg	attttacaag	agtgcataca	atacaatgta	tccaagactg	tattttctgat	61140
	tttatcgtac	cactaaactc	acaaatgtgg	ccctattctt	gtgttcacga	ctgacatcac	61200
	cgatcatggc	caagtctgat	aatagaaatg	gcattgtcac	tttcttccct	actgcaacag	61260
40	aagcccagct	atttgtctcc	cattttctct	acttctaaaa	tacatttctt	cactaagtga	61320
	gaataatctt	ttaaagacac	aatcaaaacc	atgccaccac	ctttcttgaa	ttattcaata	61380
45	tctttcgttg	gcttccagggt	tacagaaaaa	taacttgtaa	caaagtttaa	aggtcattca	61440
	tggctcctct	ctaccctatt	ttataacatt	tccccttggtg	atcagaatct	caggcacatc	61500
	atccatcttt	ctatatacaa	ataaagtcatt	atagtttgaa	ctcacctctg	gttactttta	61560
50	atcaacccaaa	tgctgtaaaa	tgcatthtga	tcgctacgtg	ttaagcagta	gttgattctt	61620
	ttcattttctg	tgtaatatcc	tattctttga	ctataccgta	atttatcaat	tctactgttg	61680
	gtaagcattt	aagtggctac	cggtttgagg	tttttatgat	tattgctgtc	ataagcattt	61740
55	ctatacatgt	ctttggatac	acacatgcat	gtgtttctga	atatctaaaa	atgtaattgc	61800
	taggtaatat	acttatcaag	catccagcat	ttgtggatac	tattaaagggt	tttccaaagg	61860

EP 3 620 521 A1

	ggttatacta ttgtacagtg tcaccaacag agtttgagtt tctattgatc catatcacca	61920
	ccaaaatttg aactgtcagt cttatctctt ctcttgtctc ttttttctc tttttttcc	61980
5	ttcccttccc ctctcttctg ttcttttctc tctcttctc ttctttctc tcttcccttc	62040
	cctttctctt tctcttccct atcccttctc ctctcctctc cctcctttt ttctcctctc	62100
	ctctccatta tttatttttc cttcttctcc tccatccctt ccatacctc tcttccctc	62160
10	ttccttctt cctttctcca tttcttctc ctctttctt caatccttc ttttgatat	62220
	gctcatgggt gtgtatttgt ctgccattgt ggcattattt gaattcagaa aagagtga	62280
15	aactactggg atcttcattc ctgggtctaa ttccacattt ttttttaaga acacatctgt	62340
	aaaaatgttc tgtactagca tattcccagg aacttcgtta aatttaatct ggctgaat	62400
	ggtaaatcta cttttcactt tgcattctt ctttagtcat accataattt taaacattca	62460
20	aaatatttgt atataatatt tgattttatc tgtcattaaa atgttaacct taaaattcat	62520
	gtttccagaa cctatttcaa taactggtaa ataaacacta ttcattttt aaatattctt	62580
	ttaatggata tttatttcaa tataataaaa aattagagtt ttattatagg aagaatttac	62640
25	caaaagaagg aggaagcaag caagttaaa ctgcagcaat agatttgtcc attccaacct	62700
	ctcaaaattc ccttggagac aaaaatctct agaggcaaag aagaacttta tattgagtca	62760
	acttgtaaaa acatctgctt ttagataagt tttcttagta taaagtgaca gaaacaaata	62820
30	agttaaactc taagatacat tccactatat tagcctaaaa cacttctgca aaaatgaaac	62880
	taggaggata ttttttagaa caactgctga aagagatgag gtggggagat atgtagagga	62940
35	gaacaggggt tctgagtcaa gacacacatg acagaacagc caatctcagg gcaagttaag	63000
	ggaatagtgg aatgaagggt catttttcat tctcaciaac taatgaaacc ctgcttatct	63060
	taaaccaacc tgctcactgg agcagggagg acaggaccag cataaaaggc agggcagagt	63120
40	cgactgttgc ttacactttc ttctgacata acagtgttca ctagcaacct caaacagaca	63180
	ccatggtgca tctgactcct gaggagaaga ctgctgtcaa tgcctgtgg ggcaaagtga	63240
	acgtggatgc agttggtggt gaggccctgg gcagggttgg atcaagggtta taagagaggc	63300
45	tcaaggaggc aaatggaaac tgggcatgtg tagacagaga agactcttgg gtttctgata	63360
	ggcactgact ctctgtccct tgggctgttt tctaccctc agattactgg tggctacctc	63420
	ttggacccag aggttctttg agtcctttgg ggatctgtcc tctcctgatg ctgttatggg	63480
50	caaccctaag gtgaaggctc atggcaagaa ggtgctaggt gccttttagtg atggcctggc	63540
	tcacctggac aacctcaagg gcactttttc tcagctgagt gagctgcact gtgacaagct	63600
55	gcacgtggat cctgagaact tcagggtgag tccaggagat gcttcacttt tctcttttta	63660
	ctttctaatac ttacattttg gttcttttac ctacctgctc ttctcccaca tttttgtcat	63720

EP 3 620 521 A1

	tttactatat	tttatcattt	aatgcttcta	aaattttgtt	aattttttat	ttaaattatc	63780
	tgcatttttt	ccttcctcac	aatccttgcta	ttttaaatta	tttaatatcc	tgtctttctc	63840
5	tcccaacccc	ctcccttcat	ttttccttct	ctaacaacaa	ctcaaattat	gcataaccagc	63900
	tctcacctgc	taattctgca	cttagaataa	tccttttgtc	tctccacatg	ggtatgggag	63960
	aggctccaac	tcaaagatga	gaggcataga	atactgtttt	agaggctata	aatcattttta	64020
10	caataaggaa	taattggaat	tttataaatt	ctgtagtaaa	tggaatggaa	aggaaagtga	64080
	atatttgatt	atgaaagact	aggcagttac	actggagggtg	gggcagaagt	cgttgctagg	64140
	agacagccca	tcatcacact	gattaatcaa	ttaatttgta	tctattaatc	tgtttatagt	64200
15	aattaatttg	tatatgctat	atacacatac	aaaattaaaa	ctaatttgga	attaatttgt	64260
	atatagtatt	atacagcata	tatagcatat	atgtacatat	atagactaca	tgctagttaa	64320
	gtacatagag	gatgtgtgtg	tatagatata	tgttatatgt	atgcattcat	atatgtactt	64380
20	atttatgctg	atgggaataa	cctggggatc	agttttgtct	aagatttggg	cagaaaaaaaa	64440
	tgggtgttgg	ctcagtttct	cagaagccag	tctttatttc	tctgttaacc	atatgcatgt	64500
	atctgcctac	ctcttctccg	cagctcttgg	gcaatgtgct	ggtgtgtgtg	ctggcccgcga	64560
25	actttggcaa	ggaattcacc	ccacaaatgc	aggctgccta	tcagaagggtg	gtggctggtg	64620
	tggctaatagc	cctggctcac	aagtaccatt	gagatcctgg	actgtttcct	gataaccata	64680
	agaagaccct	atttccctag	attctatttt	ctgaacttgg	gaacacaatg	cctacttcaa	64740
30	gggtatggct	tctgcctaata	aaagaatgtt	cagctcaact	tcctgattaa	tttcaattat	64800
	ttcatttttt	tgtccaggtg	tgtaagaagg	ttcctgaggg	tctacagata	gggagcactt	64860
35	gtttattttta	caaagagtac	atgggaaaag	agaaaagcaa	gggaaccgta	caaggcatta	64920
	atgggtgaca	cttctacctc	caaagagcag	aaattatcaa	gaactcttga	tacaaagata	64980
	atactggcac	tgcagaggtt	ctaggggaaga	cctcaaccct	aagacatagc	ctcaagggtta	65040
40	atgctacgat	taaactccaa	caattactga	gaaaataatg	tgctcaatta	aaggcataat	65100
	gattactcaa	gacaatgtta	tgttgtcttt	cttctctctt	cctttgcctg	cacattgtag	65160
	cccataatac	tataccccat	caagtgttcc	tgctccaaga	aatagcttcc	tcctcttact	65220
45	tgccccagaa	catctctgta	aagaatttcc	tcttatcttc	ccatatttca	gtcaagattc	65280
	attgctcacg	tattacttgt	gacctctctt	gaccccagcc	acaataaact	tctctatact	65340
	acccaaaaaa	tctttccaaa	ccctccccc	caccattttt	tatattttta	tatttttctt	65400
50	atttatttca	tgcacacaca	cacactccgt	gctttataag	caattctgcc	tattctctac	65460
	cttcttacat	gcctactgtg	cctcatatta	aattcatcaa	tgggcagaaa	gaaaatattt	65520
55	attcaagaaa	acagtgaatg	aatgaacgaa	tgagtaaagt	agtaaataag	ggaatgatta	65580
	ttccttgctt	tagaacttct	ggaattagag	gacaatatta	ataataccat	cgcacagtgt	65640

EP 3 620 521 A1

	ttctttgttg ttaatgctac aacatacaaa gaggaagcat gcagtaaaca accgaacagt	65700
	tatttccttt ctgatcatag gagtaatat tttttccttg agcaccattt ttgccatagg	65760
5	taaaattaga aggattttta gaactttctc agttgtatac atttttaaaa atctgtatta	65820
	tatgcatggt gattaatttt aaacttactt gaatacctaa acagaatctg ttgtttcctt	65880
	gtgtttgaaa gtgctttcac agtaactctg tctgtactgc cagaatatac tgacaatgtg	65940
10	ttatagttaa ctgttttgat cacaacattt tgaattgact ggcagcagaa gctcttttat	66000
	atccatgtgt tttccttaag tcattataca tagtaggcac tgagaactct ttatatctga	66060
15	ataagatatt taggaaccac tggtttacat atcagaagca gagctactca gggcattttg	66120
	gggaagatca ctttcacatt cctgagcata gggaagttct cataagagta agatattaaa	66180
	aggagatact tgtgtggtat tcgaaagaca gtaagagaga ttgtagacct tatgatcttg	66240
20	atagggaaaa caaactacat tcctttctcc aaaagtcaaa aaaaaagagc aaatatagct	66300
	tactataacct tctattccta caccattaga agtagtcagt gagtctaggc aagatgttgg	66360
	ccctaaaaat ccaaatacca gagaattcat gagaacatca cctggatggg acatgtgccg	66420
25	agcacacaca attactatat gctaggcatt gctatcttca tattgaagat gaggaggtca	66480
	agagatgaaa aaagacttgg caccttggtt ttatattaaa attatttgtt agagtagagc	66540
	ttttgtaaga gtctaggagt gtgggagcta aatgatgata cacatggaca caaaaaatag	66600
30	atcaacagac acccaggcct acttgagggt tgagggtggg aagagggaga cgatgaaaaa	66660
	gaacctattg ggtattaagt tcatcactga gtgatgaaat aatctgtaca tcaagacca	66720
35	gtgatatgca atttacctat ataacttgta catgtacccc caaatttaaa atgaaagtta	66780
	aaacaaagta taggaatgga attaatcct caagatttgg ctttaatttt atttgataat	66840
	ttatcaaagt gttgtttttc ttttctcact atggcggtgc tttataaact atgttcagta	66900
40	tgtctgaatg aaagggtgtg tgtgtgtgtg aaagagaggg agagaggaag ggaagagagg	66960
	acgtaataat gtgaatttga gttcatgaaa atttttcaat aaaataattt aatgtcagga	67020
	gaattaagcc taatagtctc ctaaatcatc catctcttga gcttcagagc agtcctctga	67080
45	attaatgcct acatgtttgt aaagggtgtt cagactgaag ccaagattct acctataag	67140
	agatgcaatc tcaaatttat ctgaagactg tacctctgct ctccataaat tgacaccatg	67200
	gccacttaa tgagggttaa aaaaagctaa ttctgaatga aaatctgagc ccagtggagg	67260
50	aaatattaat gaacaagggtg cagactgaaa tataaatttt tctgtaataa ttatgcatat	67320
	actttagcaa agttctgtct atgttgactt tattgctttt tggttaagaaa tacaactttt	67380
55	taaagtgaac taaactatcc tatttccaaa ctattttgtg tgtgtgcggg ttgtttctat	67440
	gggttctggt tttcttgag catttttatt tcattttaat taattaattc tgagagctgc	67500

EP 3 620 521 A1

	tgagttgtgt	ttactgagag	attgtgtatc	tgcgagagaa	gtctgtagca	agtagctaga	67560
	ctgtgcttga	cctaggaaca	tatacagtag	attgctaaaa	tgtctcactt	ggggaatttt	67620
5	agactaaaca	gtagagcatg	tataaaaaata	ctctagtcaa	gtgctgcttt	tgaacaaat	67680
	gataaaaacca	cactcccata	gatgagtgtc	atgattttca	tggaggaagt	taatattcat	67740
	cctctaagta	taccagact	agggccattc	tgatataaaa	cattaggact	taagaaagat	67800
10	taatagactg	gagtaaagga	aatggacctc	tgtctctctc	gctgtctctt	ttttgaggac	67860
	ttgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	gtcagtgggg	ctggaataaa	67920
	agtagaatag	acctgcacct	gctgtggcat	ccattcacag	agtagaagca	agctcacaat	67980
15	agtgaagatg	tcagtaagct	tgaatagttt	ttcaggaact	ttgaatgctg	atttagattt	68040
	gaaactgagg	ctctgaccat	aaccaaattt	gcactattta	ttgcttcttg	aaacttattt	68100
	gcctggtatg	cctgggcttt	tgatggctct	agtatagctt	gcagccttgt	ccctgcaggg	68160
20	tattatgggt	aatagaaaga	aaagtctgcg	ttacactcta	gtcacactaa	gtaactacca	68220
	ttggaaaagc	aaccctgcc	ttgaagccag	gatgatggta	tctgcagcag	ttgccaacac	68280
25	aagagaagga	tccatagttc	atcatttaaa	aaagaaaaca	aaatagaaaa	aggaaaacta	68340
	tttctgagca	taagaagttg	tagggtaagt	ctttaagaag	gtgacaattt	ctgccaatca	68400
	ggattttcaa	gctcttgctt	tgacaatttt	ggtctttcag	aatactataa	atataaccta	68460
30	tattataatt	tcataaagtc	tgtgcatttt	ctttgaccca	ggatatttgc	aaaagacata	68520
	ttcaaaacttc	cgcagaacac	tttatttcac	atatacatgc	ctcttatatc	agggatgtga	68580
	aacagggctc	tgaaaactgt	ctaaatctaa	aacaatgcta	atgcaggttt	aaatttaata	68640
35	aaataaaaatc	caaaatctaa	cagccaagtc	aaatctgcat	gttttaacat	ttaaaatatt	68700
	ttaaagacgt	cttttcccag	gattcaacat	gtgaaatctt	ttctcaggga	tacacgtgtg	68760
	cctagatcct	cattgcttta	gttttttaca	gaggaatgaa	tataaaaaga	aaataactta	68820
40	attttatccc	tcttacctct	ataatcatat	ataggcataa	ttttttaacc	taggctccag	68880
	atagccatag	aagaaccaa	cactttctgc	gtgtgtgaga	ataatcagag	tgagattttt	68940
	tcacaagtac	ctgatgaggg	ttgagacagg	tagaaaaagt	gagagatctc	tattttattta	69000
45	gcaataatag	agaaagcatt	taagagaata	aagcaatgga	aataagaaat	ttgtaaattt	69060
	ccttctgata	actagaaata	gaggatccag	tttcttttgg	ttaacctaaa	ttttattttca	69120
	ttttattgtt	ttattttatt	ttattttatt	ttattttgtg	taatcgtagt	ttcagagtgt	69180
50	tagagctgaa	aggaagaagt	aggagaaaca	tgcaaagtaa	aagtataaca	ctttccttac	69240
	taaaccgaca	tgggtttcca	ggtaggggca	ggattcagga	tgactgacag	ggcccttagg	69300
55	gaacactgag	accctacgct	gacctcataa	atgcttgcta	cctttgctgt	tttaattaca	69360
	tcttttaata	gcaggaagca	gaactctgca	cttcaaaagt	ttttcctcac	ctgaggagtt	69420

EP 3 620 521 A1

	aatttagtac	aaggggaaaa	agtacagggg	gatgggagaa	aggcgatcac	gttgggaagc	69480
	tatagagaaa	gaagagtaaa	tttttagtaaa	ggaggtttta	acaaacaaaa	tataaagaga	69540
5	aataggaact	tgaatcaagg	aaatgatttt	aaaacgcagt	attcttagtg	gactagagga	69600
	aaaaaataat	ctgagccaag	tagaagacct	tttccctcc	taccctact	ttctaagtca	69660
	cagaggcttt	ttgttcccc	agacactctt	gcagattagt	ccaggcagaa	acagttagat	69720
10	gtccccagtt	aacctcctat	ttgacaccac	tgattacccc	attgatagtc	acactttggg	69780
	ttgtaagtga	ctttttattt	atttgtattt	ttgactgcat	taagaggtct	ctagtttttt	69840
	atctcttggt	tcccaaaacc	taataagtaa	ctaattgcaca	gagcacattg	atttgtattt	69900
15	attctatttt	tagacataat	ttattagcat	gcatgagcaa	attaagaaaa	acaacaacaa	69960
	atgaatgcat	atatatgtat	atgtatgtgt	gtatatatac	acacatatat	atatatatatt	70020
20	tttcttttct	taccagaagg	ttttaatcca	aataaggaga	agatatgctt	agaaccgagg	70080
	tagagttttc	atccattctg	tcctgtaagt	attttgcata	ttctggagac	gcaggaagag	70140
	atccatctac	atatcccaaa	gctgaattat	ggtagacaaa	actcttccac	ttttagtgc	70200
25	tcaacttctt	atttgtgtaa	taagaaaatt	gggaaaacga	tcttcaatat	gcttaccaag	70260
	ctgtgattcc	aaatattacg	taaatacact	tgcaaaggag	gatgttttta	gtagcaattt	70320
	gtactgatgg	tatggggcca	agagatatat	cttagagggg	gggctgaggg	tttgaagtcc	70380
30	aactcctaag	ccagtgccag	aagagccaag	gacaggtagc	gctgtcatca	cttagacctc	70440
	acctgtgga	gccacacctt	agggttggcc	aatctactcc	caggagcagg	gagggcagga	70500
35	gccagggctg	ggcataaaaag	tcagggcaga	gccatctatt	gcttacattt	gcttctgaca	70560
	caactgtggt	cactagcaac	ctcaaacaga	caccatgggt	catctgactc	ctgaggagaa	70620
	gtctgcctgt	actgccctgt	ggggcaaggt	gaacgtggat	gaagttgggt	gtgaggccct	70680
40	gggcaggttg	gtatcaaggt	tacaagacag	gtttaaggag	accaatagaa	actgggcatg	70740
	tggagacaga	gaagactctt	gggtttctga	taggcactga	ctctctctgc	ctattggtct	70800
	attttcccac	ccttaggctg	ctggtggtct	accttggac	ccagagggtc	tttgagtcct	70860
45	ttggggatct	gtccactcct	gatgctgtta	tgggcaaccc	taaggtgaag	gctcatggca	70920
	agaaagtgct	cggtgccttt	agtgatggcc	tggctcacct	ggacaacctc	aagggcacct	70980
	ttgccacact	gagtgagctg	cactgtgaca	agctgcacgt	ggatcctgag	aacttcaggg	71040
50	tgagtctatg	ggacgcttga	tgttttcttt	cccttcttt	tctatggtta	agttcatgtc	71100
	ataggaaggg	gataagtaac	agggtagagt	ttagaatggg	aaacagacga	atgattgcat	71160
55	cagtgtggaa	gtctcaggat	cgttttagtt	tcttttattt	gctgttcata	acaattgttt	71220
	tcttttggtt	aattcttgct	ttcttttttt	ttcttctccg	caatttttac	tattatactt	71280

EP 3 620 521 A1

	aatgccttaa cattgtgtat aacaaaagga aatatctctg agatacatta agtaacttaa	71340
	aaaaaaactt tacacagtct gcctagtaca ttactatttg gaatatatgt gtgcttattt	71400
5	gcatattcat aatctcccta ctttattttc ttttattttt aattgataca taatcattat	71460
	acatatattat ggggttaaagt gtaatgtttt aatatgtgta cacatatattga ccaaatacagg	71520
	gtaatttttgc atttgtaatt ttaaaaaatg ctttcttctt ttaatatatact tttttgttta	71580
10	tcttattttct aatactttcc ctaatctctt tctttcaggg caataatgat acaatgtatc	71640
	atgcctcttt gcaccattct aaagaataac agtgataatt tctgggttaa ggcaatagca	71700
	atatctctgc atataaatat ttctgcatat aaattgtaac tgatgtaaga ggtttcatat	71760
15	tgctaatagc agctacaatc cagctaccat tctgctttta ttttatgggtt gggataaggc	71820
	tggattattc tgagccaag ctaggccctt ttgctaataca tgttcatacc tcttatcttc	71880
	ctcccacagc tcctgggcaa cgtgctggtc tgtgtgctgg cccatcactt tggcaaagaa	71940
20	ttcacccac cagtgcaggc tgcctatcag aaagtgggtg ctggtgtggc taatgccctg	72000
	gcccacaagt atcactaagc tcgctttctt gctgtccaat ttctattaaa ggttcctttg	72060
	ttccctaagt ccaactacta aactggggga tattatgaag ggccttgagc atctggattc	72120
25	tgcctaataa aaaacattta ttttcattgc aatgatgtat ttaaattatt tctgaatatt	72180
	ttactaaaaa ggggaatgtg gaggtcagtg catttaaaac ataaagaaat gaagagctag	72240
	ttcaaaccctt gggaaaatac actatatctt aaactccatg aaagaagggtg aggctgcaaa	72300
	cagctaatagc acattggcaa cagcccctga tgcataatgcc ttattcatcc ctcagaaaag	72360
	gattcaagta gaggcttgat ttggagggtta aagttttgct atgctgtatt ttacattact	72420
35	tattgtttta gctgtcctca tgaatgtctt ttcaactacc atttgcttat cctgcatctc	72480
	tcagccttga ctccactcag ttctcttgct tagagatacc acctttcccc tgaagtgttc	72540
	cttccatgtt ttacggcgag atggtttctc ctgcctggc cactcagcct tagttgtctc	72600
40	tgttgtctta tagaggctca cttgaagaag gaaaaacagg ggtcatgggtt tgactgtcct	72660
	gtgagccctt cttccctgcc tccccactc acagtgaacc ggaatctgca gtgctagtct	72720
	cccggaaacta tcactctttc acagtctgct ttggaaggac tgggcttagt atgaaaagtt	72780
45	aggactgaga agaatttgaa aggcggcttt ttgtagcttg atattcacta ctgtcttatt	72840
	accctgtcat aggccacccc caaatggaag tcccatctt cctcaggatg tttaagatta	72900
	gcattcagga agagatcaga ggtctgctgg ctcccttata atgtccctta tgggtgctct	72960
50	ggctctgcag ttattagcat agtgttacca tcaaccacct taacttcatt tttcttattc	73020
	aatacctagg taggtagatg ctagattctg gaaataaaat atgagtctca agtggctcctt	73080
55	gtcctctctc ccagtcaaat tctgaatcta gttggcaaga ttctgaaatc aaggcatata	73140
	atcagtaata agtgatgata gaagggtata tagaagaatt ttattatatg agagggtgaa	73200

EP 3 620 521 A1

	accctcaaaa	tgaaatgaaa	tcagaccctt	gtcttacacc	ataaacaaaa	ataaatttga	73260
	atgggttaaa	gaattaaact	aagacctaaa	accataaaaa	tttttaaaga	aatcaaaaga	73320
5	agaaaattct	aatattcacg	ttgcagccgt	tttttgaatt	tgatatgaga	agcaaaggca	73380
	acaaaaggaa	aaataaagaa	gtgaggctac	atcaaactaa	aaaatttcca	cacaaaaaac	73440
	aaaacaatga	acaaatgaaa	ggtgaaccat	gaaatggcat	atttgcaaac	caaataatttc	73500
10	ttaaatattt	tggttaatat	ccaaaatata	taagaaacac	agatgattca	ataacaaaca	73560
	aaaaattaaa	aataggaaaa	taaaaaaatt	aaaaagaaga	aaatcctgcc	atttatggca	73620
	gaattgatga	acctggagga	tgtaaaacta	agaaaaataa	gcctgacaca	aaaagacaaa	73680
15	tactacacaa	ccttgctcat	atgtgaaaca	taaaaaagtc	actctcatgg	aaacagacag	73740
	tagaggatat	gtttccaggg	gttgggggtg	ggagaatcag	gaaactatta	ctcaaagggc	73800
	ataaaatttc	agttatgtgg	gatgaataaa	ttctagatat	ctaattgtaca	gcctcgtgac	73860
20	tgtagttaat	tgtactgtaa	gtatatTTaa	aatttgcaaa	gagagtagat	tttttttttt	73920
	tttttagatg	agttttgctc	ttgttgtcca	ggctggagtg	caatggcaag	atcttggctc	73980
25	actgcaacct	ccgcctcctg	ggttcaagca	aatctcctgc	ctcagcctcc	cgagtagctg	74040
	ggattacagg	catgcgacac	catgccacgc	taattttgta	tttttagtag	agacgggggt	74100
	tctccatgtt	ggtcaggctg	atccgcctgc	ctcggccacc	caaagggctg	ggattacagg	74160
30	cgtgagccac	cgggcctggc	cgagagtaga	tcttaaaagc	atttaccaca	agaaaaaggt	74220
	aactatgtga	gataatgggt	atgttaatta	gcttgattgt	ggtaatcatt	tcacaaggta	74280
35	tacatatatt	aaaacatcat	gttgtacacc	ttaaatatat	acaattttta	tttgtgaatg	74340
	atacctcaat	aaagttgaag	aataataaaa	agaatagac	atcacatgaa	ttaaaaaact	74400
	aaaaaataaa	aaaatgcac	ttgatgatta	gaattgcatt	cttgattttt	cagatacaaa	74460
40	tatccatttg	actgtttact	cttttccaaa	acaatacaat	aaatttttagc	actttatctt	74520
	cattttcccc	ttcccaatct	ataattatat	atatatatat	tttagatatt	ttgtatagtt	74580
	ttactcccta	gatttttctag	tgttattatt	aaatagtga	gaaatgttta	cacttatgta	74640
45	caaaatgttt	tgcatgcttt	tcttcatttc	taacattctc	tctaagttta	ttctattttt	74700
	ttctgattat	ccttaatat	atctctttct	gctggaaata	cattgttact	tttggtttat	74760
	ctaaaaatgg	cttcattttc	ttcatttctaa	aatcatgtta	aattaatacc	actcatgtgt	74820
50	aagtaagata	gtggaataaa	tagaaatcca	aaaactaaat	ctcactaaaa	tataataatg	74880
	tgatatataa	aaatatagct	tttaaattta	gcttggaat	aaaaaacaaa	cagtaattga	74940
55	acaactatac	tttttgaaaa	gagtaaagtg	aaatgcttaa	ctgcatatac	cacaatcgat	75000
	tacacaatta	ggtgtgaagg	taaaattcag	tcacgaaaaa	actagaataa	aaatatggga	75060

EP 3 620 521 A1

	agacatgtat	ataatcttag	agataacact	gttatTTaat	tatcaaccca	aagtagaaac	75120
	tatcaaggga	gaaataaatt	cagtcaacaa	taaaagcatt	taagaagtta	ttctaggctg	75180
5	ggagcgggtg	ctcacacctg	caattgcagc	actttgggag	gcctagacag	gcggatcacg	75240
	acgtcaggag	ttcaagatca	gcctggccaa	catagtgaaa	cctcatcgct	actaaaaata	75300
	taaaaactta	gcctggcgctg	gtggcaggca	tgtgtaatcc	cagcaatttg	ggaggctgag	75360
10	gcaggagaat	cgcttgatcc	tgggaggcag	aggttgcagt	gagccaagat	tgtgccactg	75420
	cattccagcc	caggtgacag	catgagactc	cgtcacaaaa	aaaaaagaaa	aaaaaaaggg	75480
	gggggggagc	ggtggagcca	agatgaccga	ataggaacag	ctccagtcta	tagtctcccat	75540
15	cgtgagtgac	gcagaagacg	ggtgatttct	gcatttccaa	ctgaggtacc	aggttcatct	75600
	cacagggaag	tgccaggcag	tgggtgcagg	acagtaggtg	cagtgcactg	tgcatgagcc	75660
	aaagcagggc	gaggcatcac	ctcaccggg	aagcacaagg	ggtcagggaa	ttccctttcc	75720
20	tagtcaaaga	aaaggggtgac	agatggcacc	tggaaaatcg	ggtcactccc	gccctaatac	75780
	tgcgctcttc	caacaagctt	aacaaatggc	acaccaggag	attatatccc	atgcctggct	75840
	cagaggggtcc	tacgcccattg	gagcctcgct	cattgctagc	acagcagtct	gaggtcaaac	75900
25	tgcaaggtgg	cagtgaggct	gggggagggg	tgcccaccat	tgtccaggct	tgagcaggta	75960
	aacaaagccg	cctggaagct	cgaactgggt	ggagcccacc	acagctcaag	gaggcctgcc	76020
	tgcctctgta	ggctccacct	ctaggggcag	ggcacagaca	aacaaaagac	aacaagaacc	76080
	tctgcagact	taaatgtccc	tgtctgacag	ctttgaagag	agtagtggtt	ctcccagcac	76140
	atagcttcag	atctgagaac	aggcagactg	cctcctcaag	tgggtccctg	acccccgagt	76200
35	agcctaactg	ggaggcatcc	cccagtaggg	gcagactgac	acctcacatg	gctgggtactc	76260
	ctctaagaca	aaacttccag	aggaatgata	aggcagcagc	atttgcggtt	caccaatatc	76320
	cactgttctg	cagccaccgc	tgttgatacc	caggaaaaca	gcttctggag	tggacctcca	76380
40	gtaaaactcca	acagacctgc	agctgagggg	cctgactggt	agaaggaaaa	ctaacaaaca	76440
	gaaaggacat	ccacacccaa	aacctatctg	tacatcgcca	tcataaaga	ccaaaggtag	76500
	ataaaaccat	aaagatgggg	aaaaagcaga	gcagaaaaac	tggacactct	aaaaatgaga	76560
45	gtgcctctcc	tcctccaaag	taacgcagct	cctcaccagc	aatggaacaa	agctgggcag	76620
	agaatgactt	tgacgagttg	agagaggaag	gcttcagaag	atcaaactac	tccaagctaa	76680
	aggaggaagt	tcgaacaaac	ggcaaagaag	taaaaaactt	tgaaaaaaaa	ttagatgaat	76740
50	ggataactag	aataaccaat	gcacagaagt	ccttaaagga	cctgatggag	ctgaaaacca	76800
	aggcaggaga	actacgtgac	aaatacacaa	gcctcagtaa	ccgatgagat	caactggaag	76860
	aaaggggtatc	aatgacgaaa	gatgaaatga	atgaaatgaa	gcatgaagag	aagtttagag	76920
55	aaaaaagaat	aaaaaagaac	gaacaaagcc	tccaagaaat	atgggactat	gtgaaaagac	76980

EP 3 620 521 A1

	caaactctaca	tctaattggt	gtagctgaaa	gtgatgggga	gaatggaacc	aagtttgaaa	77040
	acactctgca	ggatattatc	caggagaact	tccccaatct	agcaaggcaa	gccccaaattc	77100
5	acattcagga	aatacagaga	acgccacaaa	gatactccta	gagaaaagca	actccaagac	77160
	acataactgt	cagattcacc	aaagttgaaa	tgaaggaaaa	aatgttaagg	gcagccagag	77220
	agaaaggctg	ggttacccac	aaagggaagc	ccatcagact	aacagctgat	ctatcggcag	77280
10	aaactctaca	agccagaaga	aagtgggggc	caatattcaa	cattgtttaa	gaaaagaatt	77340
	ttcaaccag	aatttcatat	ccagccaaac	taagcttcat	aagtgaagga	gaaataaaat	77400
	cctttacaga	caagcaaattg	ctgagagatt	ttgtcaccac	caggcctgcc	ctacaagagc	77460
15	tcctgaagga	agcactaaac	atggaaagga	acaactagta	tcagccactg	caaaaacatg	77520
	ccaaattgta	aagaccatca	aggctaggaa	gaaactgcat	caacgagcaa	aataaccagc	77580
20	taacatcata	atgacaggat	caaattcata	cataacaata	ctcaccttaa	atgtaaatag	77640
	gctaaatgct	ccaattaaaa	gacacagact	ggcaaattgg	ataaggagtc	aagacccatc	77700
	tgtgttctgt	attcaggaaa	cccatctcac	gtgcagagac	acacataggc	tcgaaataaa	77760
25	aggatggagg	aatatctacc	aagcaaattg	aaaacaaaaa	aaggcagggg	ttgcaatcct	77820
	agtctctgat	aaaacagatt	ttaaaccaac	aaagatcaaa	agagacaaag	aaggccatta	77880
	cataatggca	aagggatcta	ttcaagaaga	agaactaact	atactaaata	tatatgcacc	77940
30	caatacagga	gcacccagat	tcataaaaca	agtccctgagt	gacctacaaa	gagacttaga	78000
	tgccacacac	ataataatgg	gagactttta	cacccactg	tcaacattag	acagatcaac	78060
35	gagacagaaa	gttaacaagg	atatccagga	attggactca	gctctgcacc	aagcagacct	78120
	aatagacatc	tacagaactc	tccaccccaa	atcaacagaa	tatacattct	tttcagcacc	78180
	acaccacacc	tattccaaaa	ctgaccacat	agttggaagt	aaagctctcc	tcagcaaattg	78240
40	taaaagaaca	gaaactataa	caaactgtct	ctcagaccac	agtgcaatca	aactagaact	78300
	caggattaag	aaactcactc	aaaaccactc	agctacatgg	aaactgaaca	gcctgctcct	78360
	gaatgactac	tgggtacata	acaaaatgaa	ggcagaaata	aagatgttct	ttgaaaccaa	78420
45	cgagaacaaa	gacacaacac	accagaatct	ctgagacaca	ttcaaagcag	tgtgtagagg	78480
	gaaatttata	gcactaaatg	cccacaaggg	aaagcaggaa	agatctaaaa	ttgacaccct	78540
	aacatcacaa	ttaaaaaact	agagaagcag	gagcaaacac	attcaaaagc	taacagaaga	78600
50	caagaaataa	ctaagatcag	agcagaagtg	aaggacatag	agacacaaaa	aaacccttca	78660
	aaaaaatcaa	tgaatccaga	agctgttttt	ttgaaaagat	caacaaaatt	gatagactgc	78720
55	tagcaagact	aataaagaag	aaaagagaga	agaatcaaat	agacgcaata	aaaaatgaca	78780
	cgggggtatca	ccactgatcc	cacagaaata	caaactaccg	tcagagaata	ctataaacac	78840

EP 3 620 521 A1

	ctctacgcaa ataaactaga aaatctagaa gaaatggata aattcctcga cacatacact	78900
	ctgccaagac taaaccagga agaagttgta tctctgaata gaccaataac aggctctgaa	78960
5	attgaggcaa taattaatag cttatcaacc aaaaaaagtc cgggaccagt aggattcata	79020
	gccgaattct accagaggta caaggaggag ctggtaccat tccttctgaa actattccaa	79080
	tcaatagaaa aagaggggaat cctccctaac tcatttttatg aggccagcat catcctgata	79140
10	ccaaagcctg acagagacac aacaaaaaaa gagaatgtta caccaatatc cttgatgaac	79200
	attgatgcaa aaatcctcaa taaaatactg gcaaactgat ccaccatgat caagtgggct	79260
	tcattccctgc catgcaaggc tggttcaaca tacgaaaatc aataaacata atccagcata	79320
15	taaacagaac caaagacaca aaccatatga ttatctcaat agatgcagaa aaggcctttg	79380
	acaaaattca acaacgcttc atgctaaaaa ctctcaataa attaggtatt gatgggacat	79440
	atctcaaaat aataagagct atctatgaca aaccacacagc caatatcata ctgagtggac	79500
20	aaaaactgga agcattccct ttgaaaactg gcacaaggca gggatgccct ctctcaccac	79560
	tcctattcaa catagtgttg taagttctgg ccaggggcaat caggcaggag aaggaaataa	79620
	agggcattca attaggaaaa gaggaagtga aattgtccct gtttgcagat gacatgattg	79680
25	tatatctaga aaacccatt gtctcagccc aaaatctcct taagctgata agcaacttca	79740
	gcaaagtctc aggatataaa atcagtgtgc aaaaatcaca agtattccta tgcaccaata	79800
	acagacaaac agagagccaa atcatgagtg aactcccatt cacaattgct tcaaagagaa	79860
30	taaaatacct aggaatccaa cttacaaggg atgtgaagga cctcttcaag gagaactaca	79920
	aaccactgct caatgaaata aaagaggata caaacaaatg gaagaacatt ccatgctcat	79980
35	gggtaggaag aatcaatatc gtgaaaatgg tcatactgcc caaggttaatt tatagattca	80040
	atgccatccc catcaagcta ccaatgactt tcttcacaga actggaaaaa actactttaa	80100
	agttcatatg gaacccaaaa agagcccaca tcaccaaggc aatcctaagc caaaagaaca	80160
40	aagctggagg catcacgcta cctgacttca aactatacta caatgctacg gtaacccaaa	80220
	cagcatggta ctggtacca aacagagatc tagaccaatg gaacagaaca gagccctcag	80280
	aaataatgcc gcatatctac aactatctga tctttgacaa acctgagaga aacaagcaat	80340
45	ggggaaagga ttccctatctt aataaatggt gctgggaaaa ctggctagcc atatgtagaa	80400
	agctgaaaact ggatcccttc cttacacctt atacaaaaat taattcaaga tggattaaag	80460
	acttacatgt tagacctaaa accataaaaa ccctagaaaa aaacctaggc aataccattc	80520
50	aggacatagg catgggcaag gacttcatgt ctaaaacacc aaaagcaatg gcaacaaaag	80580
	acaaaatgga caaacgggat ctaattaaac taaagagctt ctgcacagct aaagaaacta	80640
55	ccatcagagt gaacaggcaa cctacaaaat gggagaaaaa ttttgcaatc tactcatctg	80700
	acaaagggct aatatccaga atctacaatg aactcaaaca aatttacaag aaaaaacaaa	80760

EP 3 620 521 A1

caaccccatc aaaaagtggg caaaggatat gaacagacac ttcgcaaaag aagacattta 80820

tgtaatcaaa aaacacatga aaaaatgctc atcatcacta gccatcagag aaatgcaaat 80880

5 caaaaccaca atgagatacc atctcacacc agttagaatg gcgatcatta aaaagtcagg 80940

aaacaacagg tgctggagag gatgtggaga aacaggaaca acttttacac tgttggtggg 81000

actgtaaact agttcaacca ttgcggaagt cagtgtggca attcctcagg aatctagaac 81060

10 tagaaatacc atttgacca gccatcccat tactgggtac atacccaaag gattataaat 81120

catgctgcta taaagacaca tgcacacgta tgtttattgc agcactattc acaatagcaa 81180

agacttgga ccaacccaaa tgtccaacaa cgatagactg gattaagaaa atgtggcaca 81240

15 tatacaccat ggaatactat gcagccataa aaaatgatga gttcatgtcc tttgtaggga 81300

catggatgaa gctggaaact atcattctca gcaaaactatc acaaggagaa taaaccaaac 81360

accgcatgtt ctactcata ggtgggaatt gaacaatgag aacacatgga cacatgaaga 81420

20 ggaacatcac actctgggga ctgttatggg gtggggggca ggggcaggga tagcactagg 81480

agatatacct aatgctaaat gacgagttaa tgggtgcagc acaccaacat ggcacatgta 81540

25 tacatatata acaaacctgc atgttgtgca catgtaccct aaaacttgaa gtataataat 81600

aaaaaaaagt tatcctatta aaactgatct cacacatccg tagagccatt atcaagtctt 81660

tctctttgaa atagacagaa atttagtggt ttctcagtca gttaac 81706

30

<210> 20
<211> 388
<212> DNA
<213> Homo sapiens

35

<400> 20
taagcttcag tttttcctta gttcctgtta catttctgtg tgtctccatt agtgacctcc 60

catagtccaa gcatgagcag ttctggccag gccctgtcg gggtcagtgc cccaccccg 120

40 ccttctgggt ctgtgtaacc ttctaagcaa accttctggc tcaagcacag caatgctgag 180

tcatgatgag tcatgctgag gcttaggggtg tgtgccaga tgttctcagc ctagagtgat 240

gactcctatc tgggtcccca gcaggatgct tacagggcag atggcaaaaa aaaggagaag 300

45 ctgaccacct gactaaaact ccacctcaa cggcatcata aagaaaatgg atgcctgaga 360

cagaatgtga catattctag aatatatt 388

50

<210> 21
<211> 387
<212> DNA
<213> Homo sapiens

55

<400> 21
taagcttcag tttttcctta gttcctgtta catttctgtg tgtctccatt agtgacctcc 60

EP 3 620 521 A1

	catagtccaa gcatgagcag ttctggccag gcccctgtcg gggtcagtgc cccacccccg	120
	ccttctgggt ctgtgtaacc ttctaagcaa accttctggc tcaagcacag caatgctgag	180
5	tcatgatgag tcatgctgag gctaggggtgt gtgccagat gttctcagcc tagagtgatg	240
	actcctatct ggggtccccag caggatgctt acagggcaga tggcaaaaaa aaggagaagc	300
	tgaccacctg actaaaactc cacctcaaac ggcatacataa agaaaatgga tgcctgagac	360
10	agaatgtgac atattctaga atatatt	387
	<210> 22	
	<211> 286	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 22	
	ctgagcaact aactcatgca ggactctcaa acactaacct atagcctttt ctatgtatct	60
20	acttggttag aaaccaagcg tggggactga gaaggcaata gcaggagcat tctgactctc	120
	actgcctttg gctaggtccc tccctcatca cagctcagca tagtccgagc tcttatctat	180
	atccacacac agtttctgac gctgccagc tatcaccatc ccaagtctaa agaaaaaat	240
25	aatggggttg cccatctctg ttgattagaa aacaaaacaa aataaa	286
	<210> 23	
	<211> 286	
30	<212> DNA	
	<213> Homo sapiens	
	<400> 23	
	ctgagcaact aatcatgcag gactctcaaa cactaaccta tagccttttc tatgtatcta	60
35	cttgtgtaga aaaccaagcg tgggactgag aaggcaatag caggagcatt ctgactctca	120
	ctgcctttag ctaggcccct ccctcatcac agctcagcat agtcctgagc tcttatctat	180
	atccacacac agtttctgac gctgccagc tatcaccatc ccaagtctaa agaaaaaat	240
40	aatggggttg cccatctctg ttgattagaa aacaaaacaa aataaa	286
	<210> 24	
	<211> 301	
45	<212> DNA	
	<213> Homo sapiens	
	<400> 24	
	ccaatcgtgg catatcctct aaactttctt ttcccttcat aaatcctctt tctttttttt	60
50	ccccctcaca gttttcctga acagggtgac tattaattgt gtctgcttga tgtggacacc	120
	aggtggcgct ggacatcaga tttggagagg cagttgtcta ggggaaccggg ctctgtgcca	180
	gcgcaggagg caggctggct ctctatttcc aggatgctc atccaggaag gaaaggttgc	240
55	atgctggaca cactaacctt gaagaattct tctgtctctc tcgtcattta gaaaggaagg	300

a 301

<210> 25
 <211> 305
 <212> DNA
 <213> Homo sapiens

<400> 25
 ctagccaatc gtggcatatc ctctaaactt tcttttccct tcataaatcc tctttctttt 60
 ttttccccct cacagttttc ctgaacaggt tgactattaa ttgtgtctgc ttgatgtgga 120
 caccaggtgg cgctggacat cagatttggg gaggcagttg tctaggggaa cgggctctgt 180
 gccagcgcag gaggcaggct ggctctccta ctccagggat gctcatccag gaaggaaagg 240
 ttgcatgctg gacacactaa ccttgaagaa ttcttctgtc tctctcgtca tttagaaagg 300
 aagga 305

<210> 26
 <211> 896
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 26
 ctcgagttaa ttaatctccc acgccctggg ctccagcttg ggagtgggtca gaccccaatg 60
 gcgataaact ctggcaactt tatctgtgca ctgcaggctc agccccaaca gcttttagctt 120
 tcacaagcag gcaggggaag ggaaacacat atctccagat atgagggttaa ttaacctgca 180
 ggctaaaccc ctccccacc ctagcccaa gcttcatctt agctccactc ctgacctat 240
 ccagctaaag gtccccaccc agctcctgcc tatctagtca ttgcatatgg caagacttga 300
 aagtcctatc tcaaagcagc agaattatca gctacgactc ctgcagggtta taaccatccc 360
 ccagcactcc ctgccccac agccagact tgaccaactc ccagctccgc ctgggacttc 420
 cagatatggg gccccaccct tgcaggcctt ggggacgctg aagatattga ctatctgcgt 480
 gccggaaaag ggtgttataa accggtaaag gctgggggtg ggagtagcgg atttgaagca 540
 cttgttggcc tacagaggtg tggcaagcag agcacctcag aactcaggcg tactgcccgc 600
 cgcccgagcc ctgcgagggc cgatagcgag ggtgtggccc ttatctgcac ccagcagagc 660
 gccggcgggg tacggtcacc ggtcccgggc agttgcctca gctgagtatg tcttctaaag 720
 ataatgtcga ttgtgtatgg ctgatgggat tctaggacca agcaagaggt tttttttttt 780
 cccccacata cttaacgttt ctatatctct atttgaattc gactggacag ttccatttga 840
 attatttctc tctctctctc tctctgacac attttatctt gccacccggg ctcgag 896

Claims

1. An insulator comprising the CTCF binding site sequence set forth in SEQ ID NO:18, optionally wherein the insulator comprises the nucleotide sequence set forth in SEQ ID NO:1, SEQ ID NO: 24 or SEQ ID NO:25
2. An expression cassette comprising the insulator of claim 1, and a globin gene or a functional portion thereof operably linked to a β -globin locus control region (LCR) region.
3. The expression cassette of claim 2, wherein the β -globin LCR region
 - a. does not comprise a Dnase I hypersensitive site-2 (HS2) region, and optionally
 - i. the β -globin LCR region does not comprise a core sequence of HS2; and/or
 - ii. the β -globin LCR region does not comprise a HS2 region that sustains the enhancer activity of HS2;
 - b. comprises a Dnase I hypersensitive site-1 (HS1) region, a Dnase I hypersensitive site-3 (HS3) region, and a Dnase I hypersensitive site-4 (HS4) region, and optionally the HS3 region is positioned between the HS1 region and the HS4 region;
 - c. does not comprise a HS1 region, and optionally
 - i. the β -globin LCR region does not comprise a core sequence of HS1;
 - ii. the β -globin LCR region does not comprise a HS1 region that sustains the function of HS1; and/or
 - iii. the β -globin LCR region comprises a HS3 region and a HS4 region, and optionally the HS3 region is positioned between the globin gene or functional portion thereof and the HS4 region; and/or
 - d. comprises a HS2 region, a HS3 region, and a HS4 region, and optionally further comprising a HS1 region.
4. The expression cassette of claim 3, wherein
 - (a) the core sequence of HS2 has the nucleotide sequence set forth in SEQ ID NO:20 or SEQ ID NO: 21; and/or
 - (b) the HS1 region is about 1.1 kb in length, about 600 bp in length, or about 490 bp in length, optionally the HS 1 region has the nucleotide sequence set forth in SEQ ID NO:2, SEQ ID NO:3 or SEQ ID NO:4; and/or
 - (c) the core sequence of HS1 has the nucleotide sequence set forth in SEQ ID NO:22 or SEQ ID NO:23; and/or
 - (d) the HS3 region is about 1300 bp in length, optionally the HS3 region has the nucleotide sequence set forth in SEQ ID NO:5; and/or
 - (e) the HS4 region is about 1.1 kb in length or about 450 bp in length, optionally the HS4 region has the nucleotide sequence set forth in SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8.
5. The expression cassette of claim 3, wherein
 - a. the β -globin LCR region comprises a HS1 region, a HS3 region, and a HS4 region and does not comprise a HS2 region, wherein
 - i. the HS1 region comprises the nucleotide sequence set forth in SEQ ID NO:2, the HS3 region comprises the nucleotide sequence set forth in SEQ ID NO:5, and the HS4 region comprises the nucleotide sequence set forth in SEQ ID NO:6;
 - ii. the HS1 region comprises the nucleotide sequence set forth in SEQ ID NO:3, the HS3 region comprises the nucleotide sequence set forth in SEQ ID NO:5, and the HS4 region comprises the nucleotide sequence set forth in SEQ ID NO:8; or
 - iii. the HS1 region comprises the nucleotide sequence set forth in SEQ ID NO:4, the HS3 region comprises the nucleotide sequence set forth in SEQ ID NO:5, and the HS4 region comprises the nucleotide sequence set forth in SEQ ID NO:8; or
 - b. the β -globin LCR region comprises a HS3 region and a HS4 region and does not comprise a HS1 region or a HS2 region, wherein the HS3 region comprises the nucleotide sequence set forth in SEQ ID NO:5 and the HS4 region comprises the nucleotide sequence set forth in SEQ ID NO:6; or
 - c. the β -globin LCR region comprises a HS2 region, a HS3 region, and a HS4 region, and the HS2 region comprises the nucleotide sequence set forth in SEQ ID NO:9, the HS3 region comprises the nucleotide sequence

set forth in SEQ ID NO:5, and the HS4 region comprises the nucleotide sequence set forth in SEQ ID NO:7; or
 d. the β -globin LCR region comprises a HS2 region, a HS3 region, and a HS4 region and the HS2 region is
 860bp in length, and/or the HS3 region is about 1300 bp in length, and/or the HS4 region is about 1.1 kb in
 length; optionally the HS2 region comprises the nucleotide sequence set forth in SEQ ID NO:9, and/or the HS3
 region comprises the nucleotide sequence set forth in SEQ ID NO:5, and/or the HS4 region comprises the
 nucleotide sequence set forth in SEQ ID NO:7.

6. The expression cassette of any one of claims 2-5, wherein the globin gene is selected from the group consisting of
 β -globin gene, γ -globin gene, and δ -globin gene, optionally the globin gene is human β -globin gene, and optionally
 the human β -globin gene is selected from the group consisting of a wild-type human β -globin gene, a deleted human
 β -globin gene comprising one or more deletions of intron sequences, and a mutated human β -globin gene encoding
 at least one anti-sickling amino acid residue, optionally the human β -globin gene is human β^A -globin gene encoding
 a threonine to glutamine mutation at codon 87 (β^{A-T87Q}).

7. The expression cassette of any one of claims 2-6, further comprising:

a. a β -globin promoter, and optionally the β -globin promoter i) is positioned between the globin gene or functional
 portion thereof and the β -globin LCR region, and/or ii) is a human β -globin promoter that is about 610 bp or
 about 265 in length, optionally the β -globin promoter has the nucleotide sequence set forth in SEQ ID NO:10
 or SEQ ID NO: 11; and/or

b. a human β -globin 3' enhancer, and optionally the human β -globin 3' enhancer i) is positioned in the upstream
 of the globin gene or functional portion thereof, and/or ii) is about 880 bp in length, and optionally the human
 β -globin 3' enhancer has the nucleotide sequence set forth in SEQ ID NO: 12; and/or

c. at least one (e.g., one, two or three) erythroid-specific enhancer, and optionally the at least one erythroid
 -specific enhancer i) is positioned between the globin gene or functional portion thereof and β -globin LCR region,
 and/or ii) the at least one erythroid-specific enhancer has a nucleotide sequence set forth in SEQ ID NO: 13,
 SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17.

8. A recombinant vector comprising the expression cassette of any one of claims 2-7, and optionally wherein the
 recombinant vector is a retroviral vector, and optionally wherein the recombinant vector is a lentivirus vector.

9. The recombinant vector of claim 8, further comprising: a Woodchuck hepatitis post-regulatory element (WPRE) in
 the 3' long terminal repeat (LTR) of the vector, and/or a bovine growth hormone polyadenylation signal in the 3' long
 terminal repeat (LTR) of the vector.

10. A targeted delivery system comprising a non-naturally occurring or engineered nuclease and a polypeptide encoded
 by the expression cassette of any one of claims 2-7, optionally the nuclease is selected from the group consisting
 of a non-naturally occurring or engineered zinc-finger nuclease (ZFN), a non-naturally occurring or engineered
 meganuclease, and a non-naturally occurring or engineered transcription activator-like effector nuclease (TALEN),
 and/or the nuclease comprises a DNA binding domain and a nuclease cleavage domain, and/or the nuclease binds
 to a genomic safe harbor site, and optionally

a. the nuclease generates a double strand break (DSB) at the genomic safe harbor site; and/or

b. the genomic safe harbor site is an extragenic genomic safe harbor site; and/or

c. the genomic safe harbor site is located on chromosome 1; and/or

d. the genomic safe harbor site meets all of the following five criteria: (1) distance of at least 50 kb from the 5'
 end of any gene (e.g., from the 5' end of the gene), (ii) distance of at least 300 kb from any cancer-related gene,
 (iii) within an open/accessible chromatin structure (measured by DNA cleavage with natural or engineered
 nucleases), (iv) location outside a gene transcription unit and (v) location outside ultraconserved regions (UCRs),
 microRNA or long non-coding RNA of the human genome.

11. A targeted delivery system comprising a non-naturally occurring or engineered CRISPR-Cas system and a polypep-
 tide encoded by the expression cassette of any one of claims 2-7, optionally the CRISPR-Cas system comprises a
 CRISPR-Cas nuclease and a single-guide RNA, and optionally the CRISPR-Cas system binds to a genomic safe
 harbor site.

12. A vector comprising a polynucleotide encoding a non-naturally occurring or engineered nuclease and the expression
 cassette of any one of claims 2-7, optionally the vector is a lentiviral vector, and optionally the nuclease is selected

from the group consisting of a non-naturally occurring or engineered zinc-finger nuclease (ZFN), a non-naturally occurring or engineered meganuclease, and a non-naturally occurring or engineered transcription activator-like effector nuclease (TALEN), and/or the nuclease comprises a DNA binding domain and a nuclease cleavage domain, and/or the nuclease binds to a genomic safe harbor site, and optionally

- a. the nuclease generates a double strand break (DSB) at the genomic safe harbor site; and/or
- b. the genomic safe harbor site is an extragenic genomic safe harbor site; and/or
- c. the genomic safe harbor site is located on chromosome 1; and/or
- d. the genomic safe harbor site meets all of the following five criteria: (i) distance of at least 50 kb from the 5' end of any gene (e.g., from the 5' end of the gene), (ii) distance of at least 300 kb from any cancer-related gene, (iii) within an open/accessible chromatin structure (measured by DNA cleavage with natural or engineered nucleases), (iv) location outside a gene transcription unit and (v) location outside ultraconserved regions (UCRs), microRNA or long non-coding RNA of the human genome.

13. A vector comprising a polynucleotide encoding a non-naturally occurring or engineered CRISPR-Cas system and the expression cassette of any one of claims 2-7, optionally the CRISPR-Cas system comprises a CRISPR-Cas nuclease and a single-guide RNA, and optionally the CRISPR-Cas system binds to a genomic safe harbor site, and optionally the vector is a lentiviral vector.

14. A cell transduced with the expression cassette of any one of claims 2-7, the recombinant vector of claim 8 or 9, the targeted delivery system of claim 10 or 11, or the vector of claim 12 or 13, optionally the cell is selected from the group consisting of a hematopoietic stem cell, an embryonic stem cell, an induced pluripotent stem cell, and a hemogenic endothelium cell, and optionally the cell is transduced *ex vivo*, and optionally the hematopoietic stem cell is a CD34⁺ hematopoietic stem cell.

15. A composition comprising the cell of claim 14, optionally the composition is a pharmaceutical composition further comprising a pharmaceutically acceptable.

16. The cell of claim 14 or the composition for claim 15 for use in treating a hemoglobinopathy in a subject or for use in restoring a subject's ability to produce red blood cells containing normal hemoglobin, optionally wherein the hemoglobinopathy is selected from the group consisting of hemoglobin C disease, hemoglobin sickle cell disease (SCD), sickle cell anemia, hereditary anemia, thalassemia, β -thalassemia, thalassemia major, thalassemia intermedia, α -thalassemia, and hemoglobin H disease.

17. A kit for treating a hemoglobinopathy comprising the cell of claim 14, and optionally further comprising written instructions for using the cell for treating a subject having a hemoglobinopathy, and/or wherein the hemoglobinopathy is selected from the group consisting of hemoglobin C disease, hemoglobin sickle cell disease (SCD), sickle cell anemia, hereditary anemia, thalassemia, β -thalassemia, thalassemia major, thalassemia intermedia, α -thalassemia, and hemoglobin H disease.

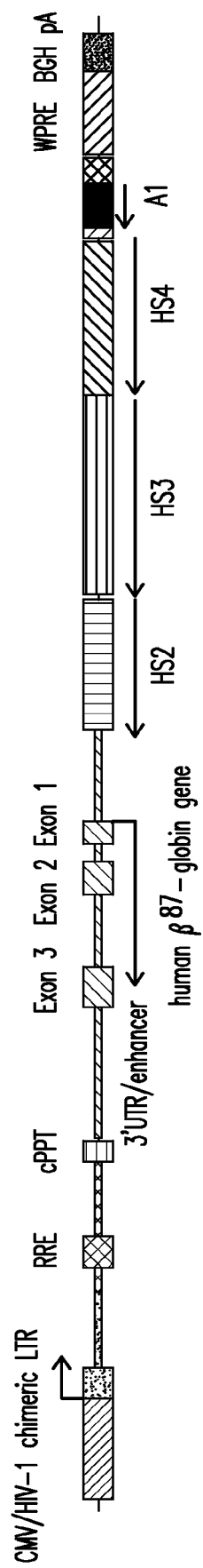


FIG. 1

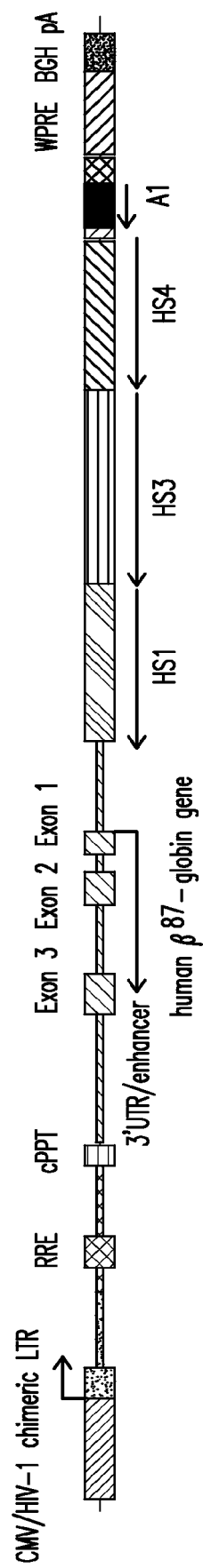


FIG. 2

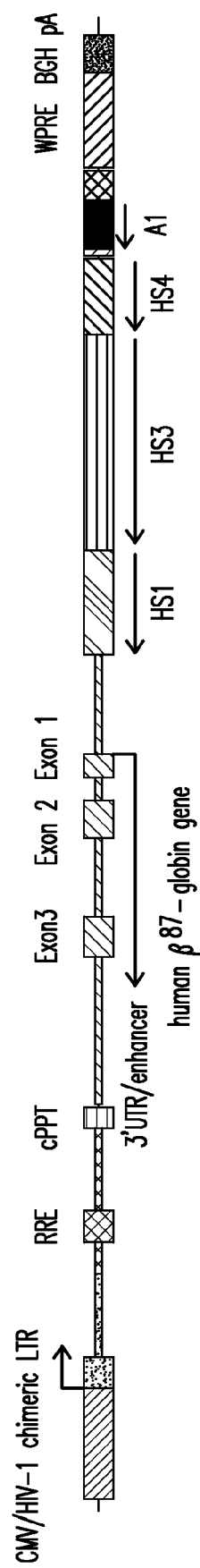


FIG. 3

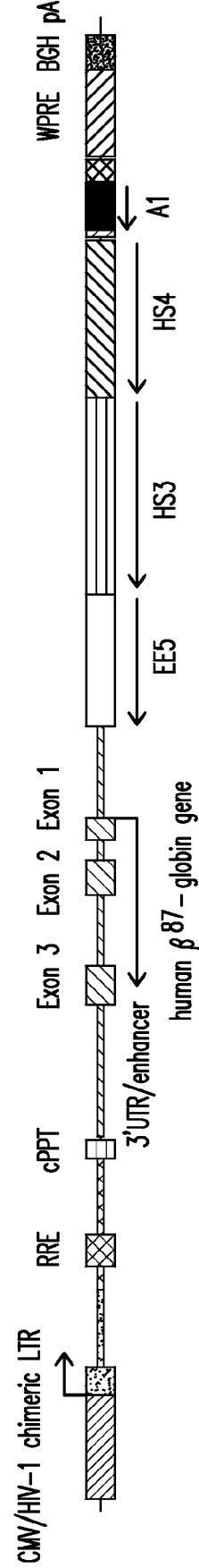
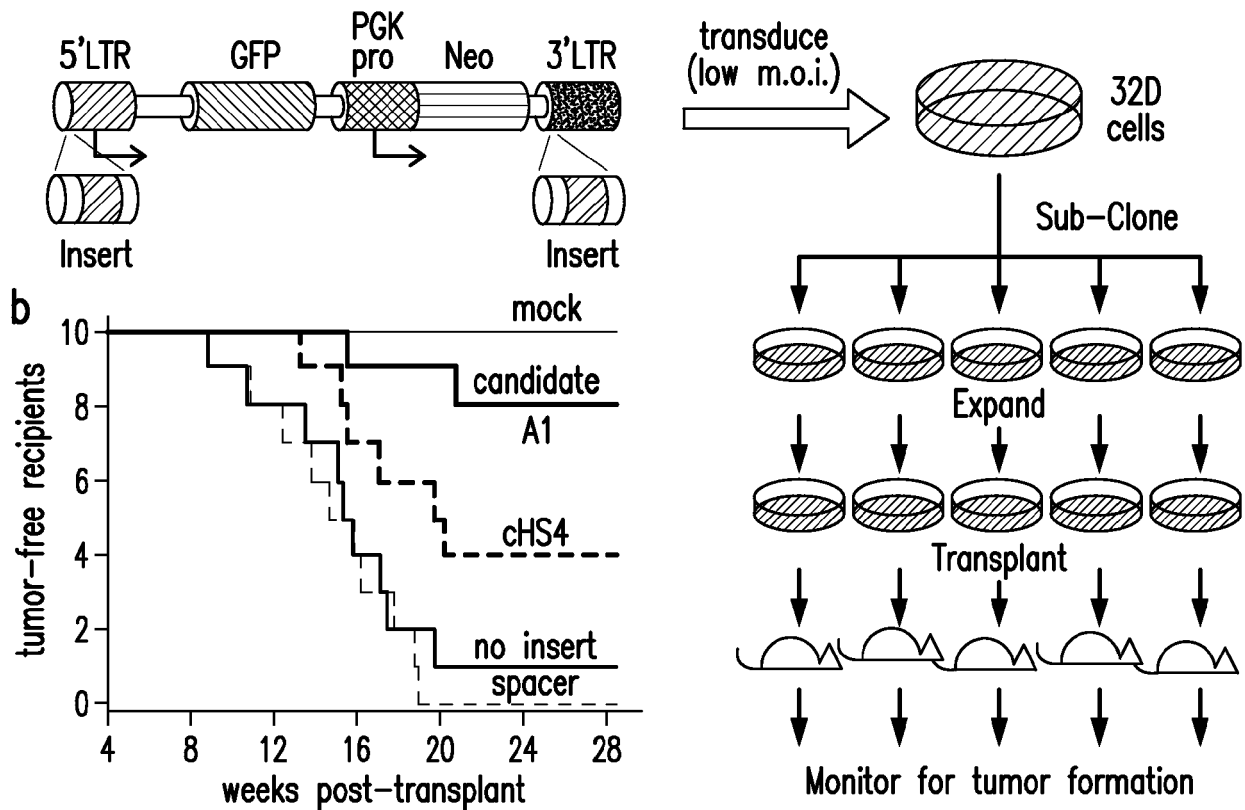


FIG. 4

a Gammaretroviral Vector Genotoxicity Assay**c**

Vector	(a) Estimated no. Tumors	(b) Estimated no. Provirus	Tumors per 10 ⁵ Provirs	(c) Probability vs. no insert	(c) Probability vs. cHS4
mock	0	0	0	<< 0.001	<< 0.001
no insert	23.0	0.49 x 10 ⁵	46.9	--	<0.001
spacer	≥30	0.67 x 10 ⁵	≥44.7	n.e.	<0.001
cHS4	9.15	0.54 x 10 ⁵	16.9	<0.001	--
A1	2.23	0.64 x 10 ⁵	3.9	<0.001	<0.05

(a) Based on the poisson distribution for the fraction of recipient with no tumores; (b) Based on the initial cell numbers and initial transduction rates; (c) Based on the *Z-test* for two proportions.

FIG. 5

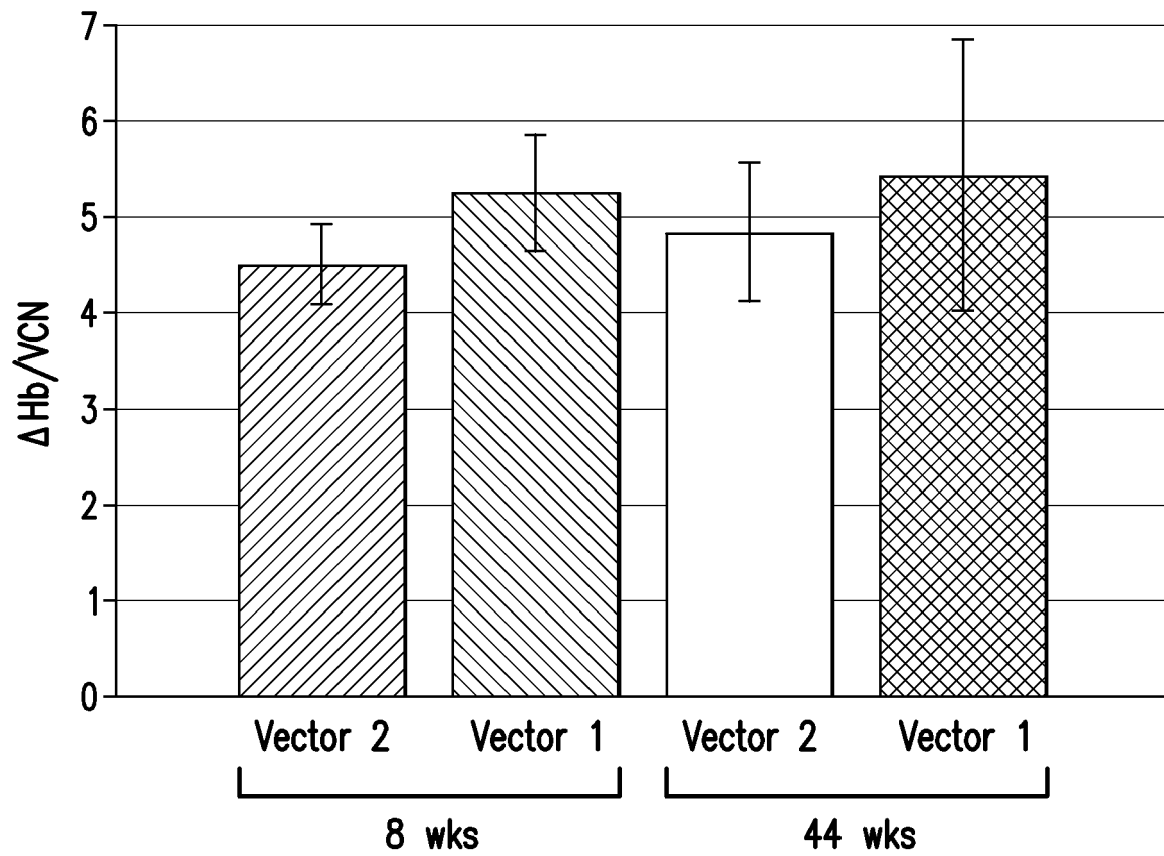


FIG. 6

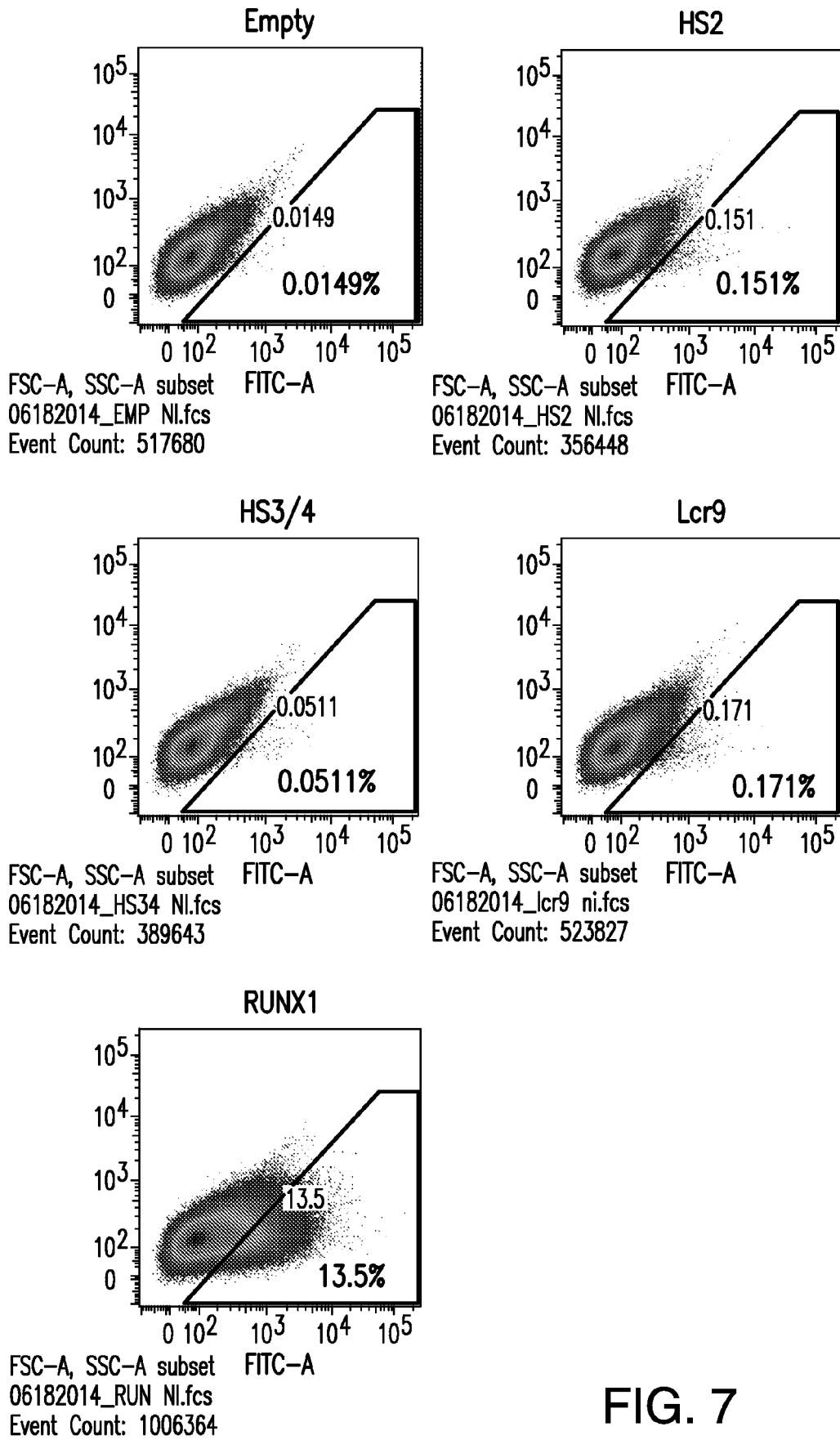
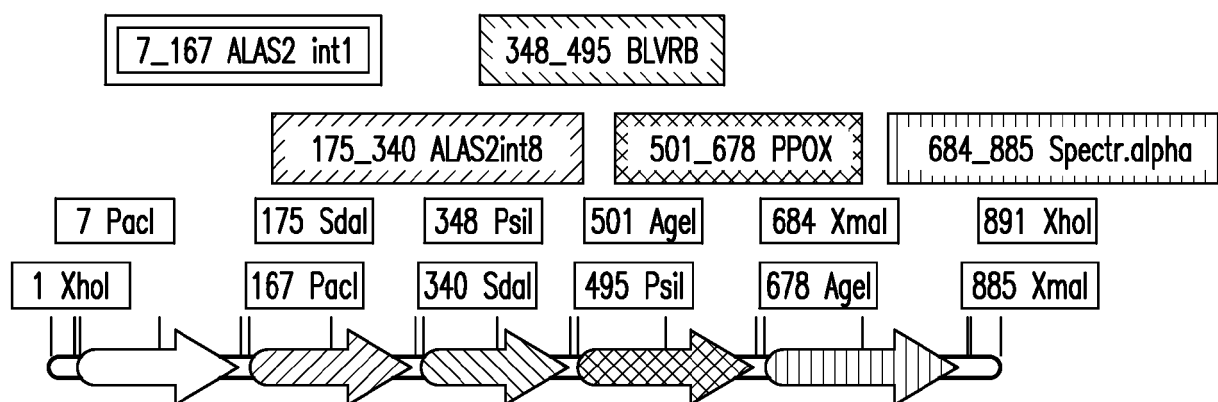


FIG. 7

- ALAS Intron 1:
 - **XhoI_PacI**~ TCTCCACGCCCTGGTCTCAGCTTGGGAGTGGTCAGACCCCAATGGGATAAACTCTGGCAACTTTATCTGTGcaCTGCAGGCTCAGCCCCA
 AcaGCTTTAGCTTTCACAAGCAGGCAGGGGAAGGGAAACACATATCTCCAGATATGAGG –PacI (TTAAT/TAA)
 - SdaI~ CTAAACCCCTCCCCACCCCTAGCCCCAAGCTTCATCTTAGCTCCACTCCTGACCCCTATCCAGCTAAAGGTCCCCACCCAGCTCCTGCCTATCTAGTCAT
 TGCATATGGCAAGACTTGAAGTCCTATCTCAAGCAGCAGAAATTATCAGCTAGCACT –SdaI (CCTGCA/GG)
 - BLVRB:
 - **PaiI**~ CCATCCCCCAGCACTCCCTGCCCCCAGCCAGACTTGACCAACTCCCAGCTccGCCCTGGGACTTCCAGATATGGGGCCCCACCCCTTGCAGGCCCTTGG
 GGACGCTGAAGATATTGACTATCTCGCTGCCggAAAAGGGTG –PaiI (TTA|TAA)
 - **PPOX**
 - **Agel**_ AAAGGCTGGGGTGGGAGTAGCGGATTTGAAGCACTTGTTGGCCTACAGAGGTGTGGCAAGCAGAGCACCCTCAGAACTCAGGCGTACTGCCCCCGCCCC
 GAGCCCTGCGAGGGCCGATAGCGAGGGTGTGGCCCTTATCTGCACCCAGCAGAGCGCCCGGGGTACGGTC –Agel (a/ccggt)
 - **Spectrin- α pha**
 - **Xma_CAG**TTGCCTCAGCTGAGTATGTCTTCTAAGATAATGTCGATTGTCGTATGGCTGATGGGATTCTAGGACCAAGCAAGAGGTTTTTTTTTCCCCACATACTTA
 ACGTTTCTATATTCTATTGAAATTCGACTGGACAGTCCATTGAAATTATTTCTCTCTCTCTCTCTCTGACACATTTTATCTTGCCA –Xma (c/ccggg) –**XhoI**

FIG. 8



new enhancer – 896 nt

• New-ENHANCER

• CTCGAGttaattaaTCTCCCACGCCCTGGTCTCAGCTTGGGGAGTGGTCAGACCCCAATGGCGATAA
 ACTCTGGCAACTTTATCTGTGcaCTGCAGGCTCAGCCCCAAcaGCTTTAGCTTTACAAGCAGGCAGGGG
 AAGGGAAACACATATCTCCAGATATGAGGttaattaacctgcaggCTAAACCCCTCCCCACCCTAGCCC
 CAAGCTTCATCTTAGCTCCACTCCTGACCCTATCCAGCTAAAGGTCCCACCCAGCTCCTGCCTATCTAG
 TCATTGCATATGGCAAGACTTGAAAGTCCTATCTCAAAGCAGCAGAATTATCAGCTACGACTcctgcagg
 ttataaCCATCCCCCAGCACTCCCTGCCCCACAGCCCAGACTTGACCAACTCCCAGCTccGCCTGGGAC
 TTCCAGATATGGGGCCCCACCCTTGAGGCCTTGGGGACGCTGAAGATATTGACTATCTGCGTGCCgga
 AAGGGTGttataaacggtaAAGGCTGGGGGTGGGAGTAGCGGATTTGAAGCACTTGTTGGCCTACAGAG
 GTGTGGCAAGCAGAGCACCTCAGAACTCAGGCGTACTGCCCGCCGCCGAGCCCTGCGAGGGCCGATAGC
 GAGGGTGTGGCCCTTATCTGCACCCAGCAGAGCGCCGGCGGGGTACGGTCaccggtcccgggCAGTTGCC
 TCAGCTGAGTATGTCTTCTAAAGATAATGTCGATTGTGTATGGCTGATGGGATTCTAGGACCAAGCAAGA
 GGTTTTTTTTTTCCCCACATACTTAACGTTTCTATATTTCTATTTGAATTCGACTGGACAGTTCCATT
 TGAATTATTTCTCTCTCTCTCTCTCTGACACATTTTATCTTGCCAcccgggCTCGAG

FIG. 9

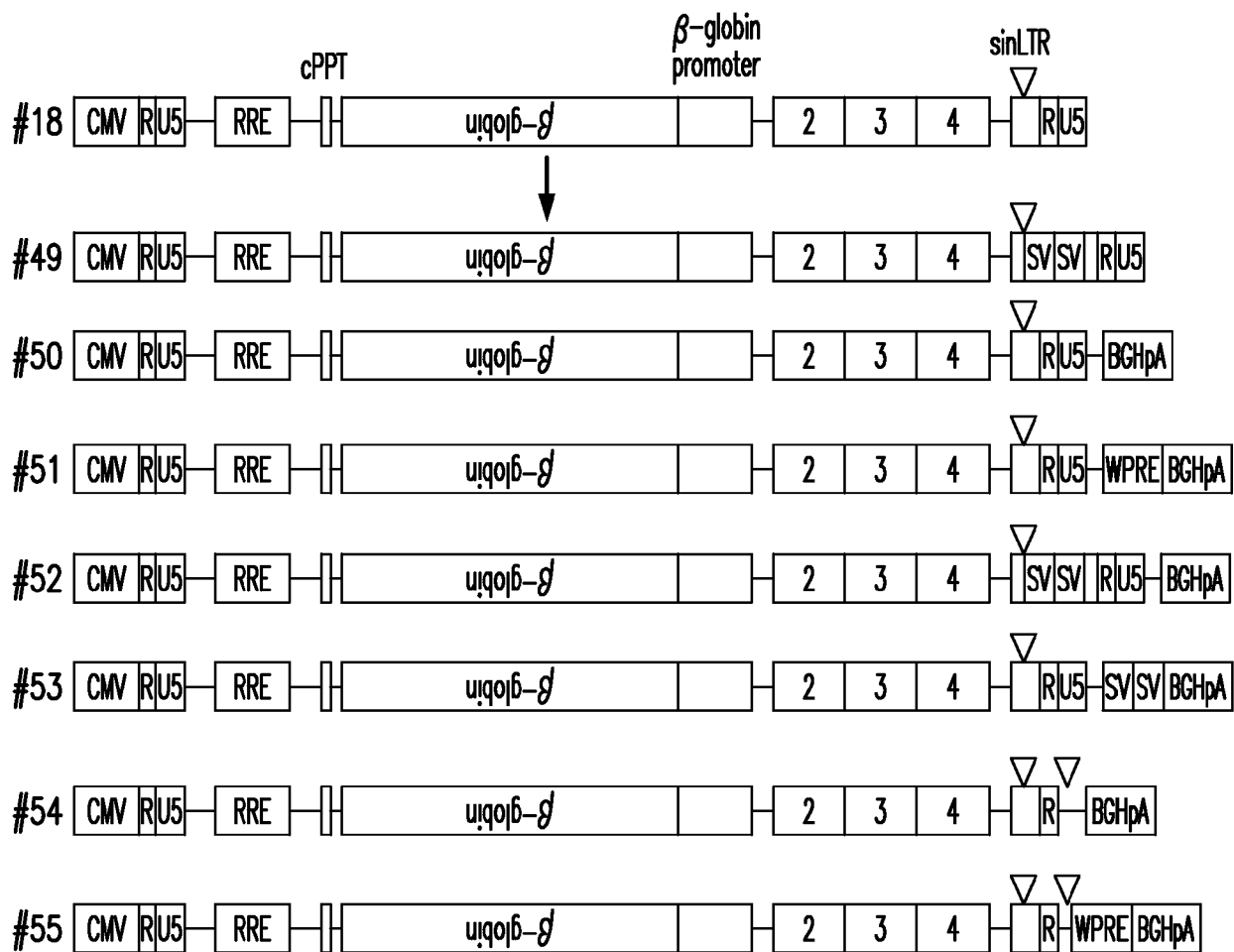


FIG. 10A

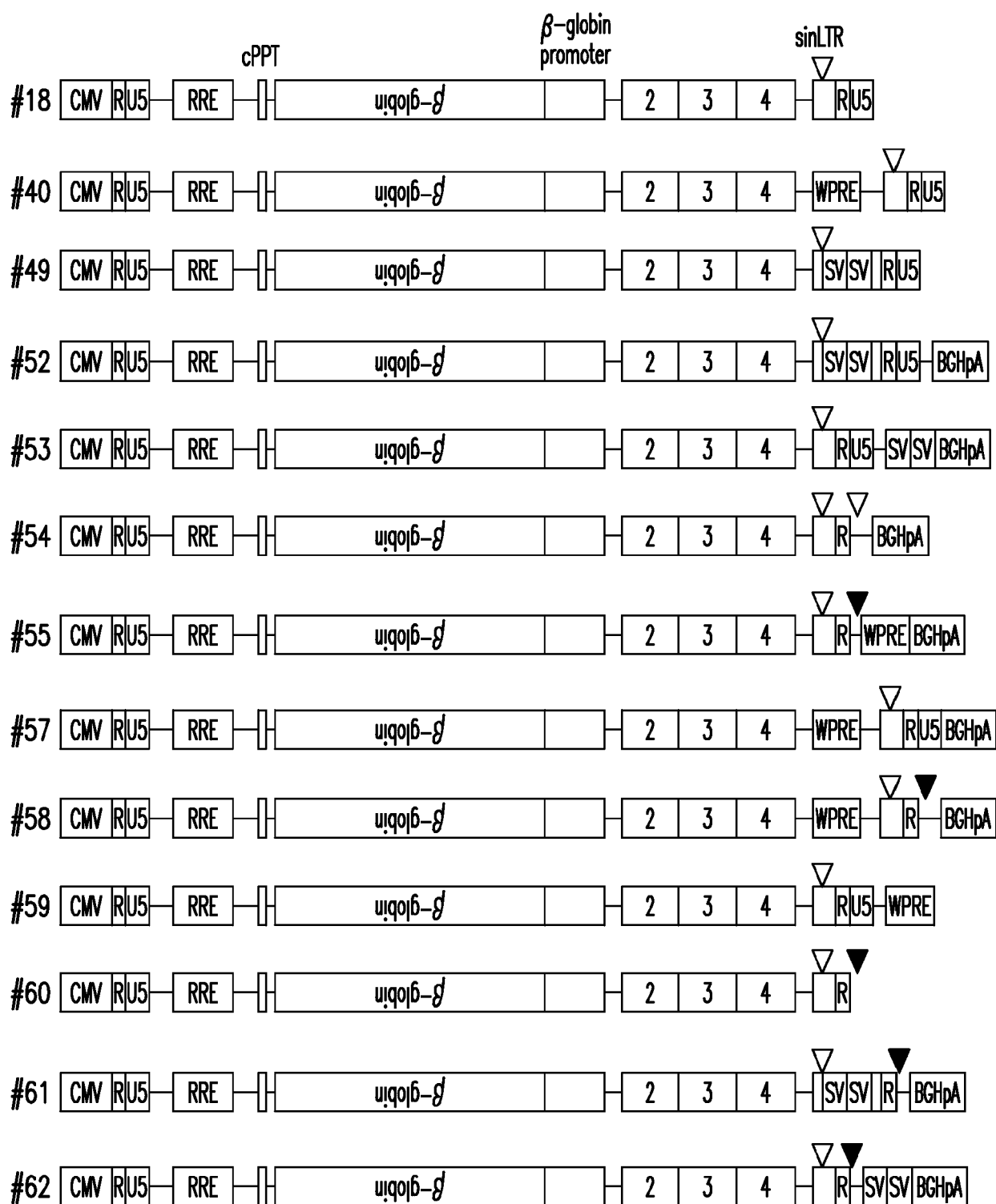


FIG. 10B

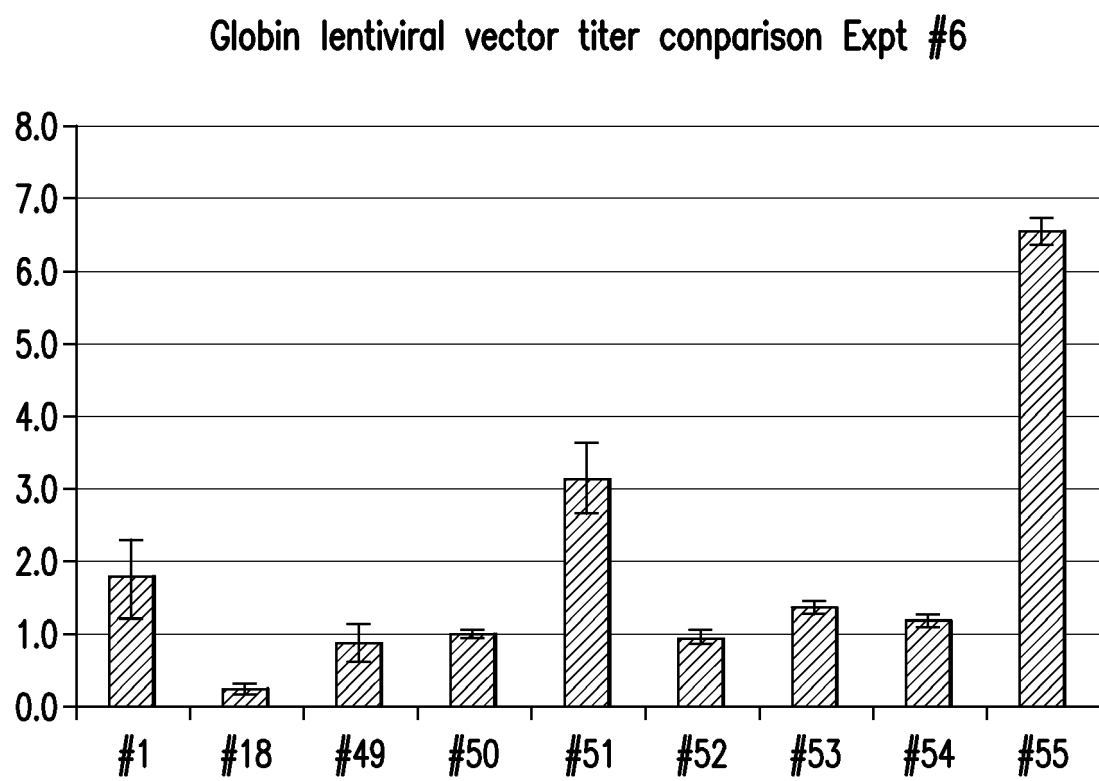


FIG. 11

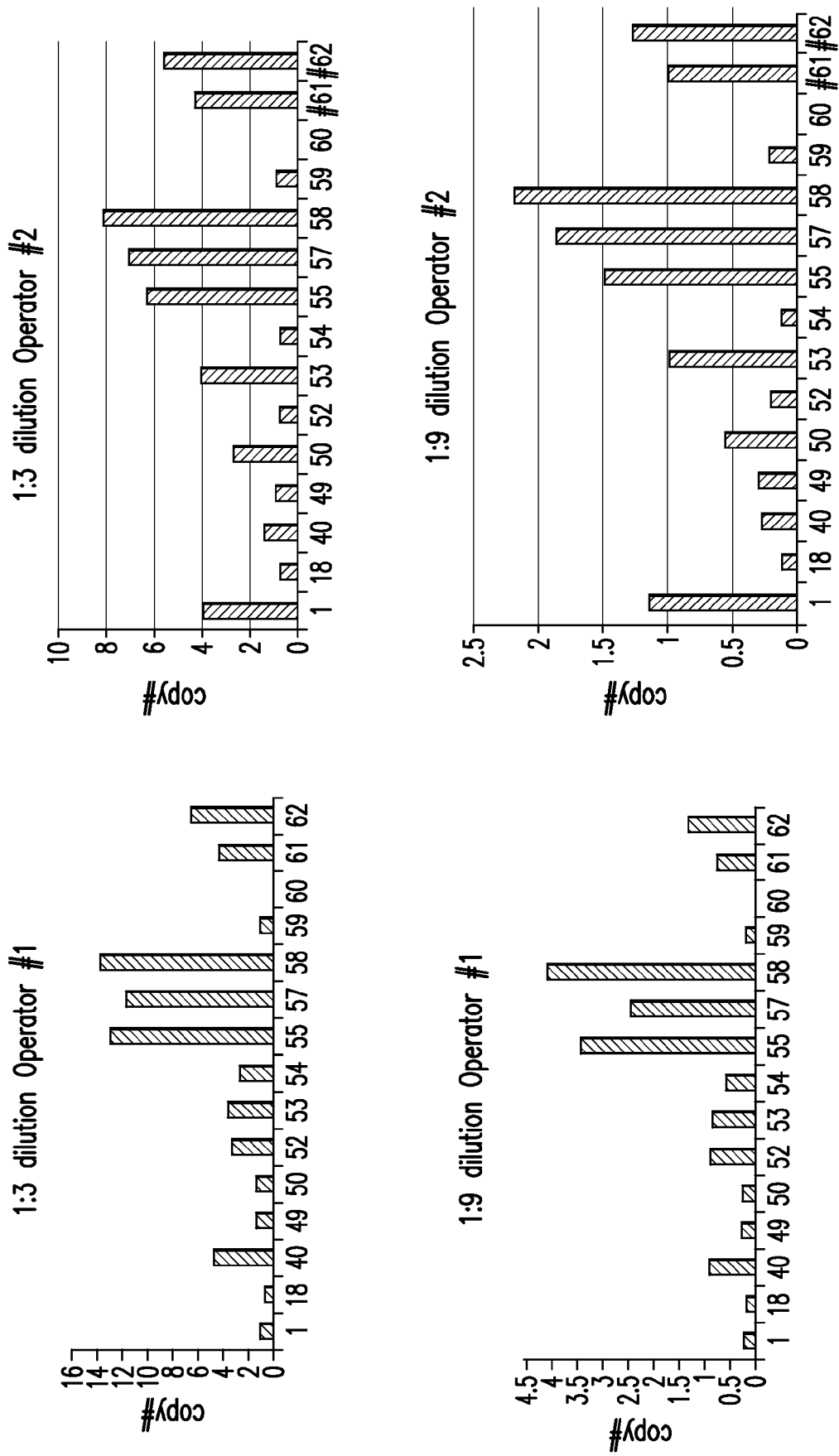


FIG. 12



EUROPEAN SEARCH REPORT

 Application Number
 EP 19 18 5730

5

10

15

20

25

30

35

40

45

50

55

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
E	WO 2015/138852 A1 (UNIV WASHINGTON [US]) 17 September 2015 (2015-09-17) * the whole document *	1	INV. C12N15/113 A61K48/00 C12N15/85 C12N15/867
X	----- DATABASE EMBL [Online] 31 July 2001 (2001-07-31), "Homo sapiens chromosome 1 clone RP11-550H2, complete sequence.", XP002777813, retrieved from EBI accession no. EM STD:AC092813 Database accession no. AC092813 * sequence *	1	
A	----- WO 2014/043131 A1 (UNIV CALIFORNIA [US]; KOHN DONALD B [US]; URBINATI FABRIZIA [US]; GARC) 20 March 2014 (2014-03-20) * figures 3,4,11 *	2-17	
A	----- WO 2010/046493 A2 (UNIV LAUSANNE [CH]; MERMOD NICOLAS [CH]; GAUSSIN ARMELLE [CH]; ESNAULT) 29 April 2010 (2010-04-29) * figure 23B *	2-17	
A	----- DAVID W. EMERY: "The Use of Chromatin Insulators to Improve the Expression and Safety of Integrating Gene Transfer Vectors", HUMAN GENE THERAPY, vol. 22, no. 6, 1 June 2011 (2011-06-01), pages 761-774, XP055061191, ISSN: 1043-0342, DOI: 10.1089/hum.2010.233 * page 763, column 2, paragraph 2 *	2-17	TECHNICAL FIELDS SEARCHED (IPC) A61K C12N A01K C07K
----- -/--			
The present search report has been drawn up for all claims			
Place of search The Hague		Date of completion of the search 28 January 2020	Examiner Lonnoy, Olivier
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document			

EPO FORM 1503 03.82 (P04C01)



EUROPEAN SEARCH REPORT

Application Number
EP 19 18 5730

5

10

15

20

25

30

35

40

45

50

55

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
A	AMY C. GROTH ET AL: "Identification and Characterization of Enhancer-Blocking Insulators to Reduce Retroviral Vector Genotoxicity", PLOS ONE, vol. 8, no. 10, 3 October 2013 (2013-10-03), page e76528, XP055447014, DOI: 10.1371/journal.pone.0076528 * figures 7-9; tables 1,2 *	2-17	
A,P	MINGDONG LIU ET AL: "Genomic discovery of potent chromatin insulators for human gene therapy", NATURE BIOTECHNOLOGY (ADVANCE ONLINE PUBLICATION), vol. 33, no. 2, 12 January 2015 (2015-01-12), pages 198-203, XP055447005, ISSN: 1087-0156, DOI: 10.1038/nbt.3062 * the whole document *	2-17	
The present search report has been drawn up for all claims			TECHNICAL FIELDS SEARCHED (IPC)
Place of search The Hague		Date of completion of the search 28 January 2020	Examiner Lonnoy, Olivier
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document			

EPO FORM 1503 03.82 (P04C01)

**ANNEX TO THE EUROPEAN SEARCH REPORT
ON EUROPEAN PATENT APPLICATION NO.**

EP 19 18 5730

5 This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

28-01-2020

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2015138852 A1	17-09-2015	AU 2015229227 A1	29-09-2016
		EP 3117004 A1	18-01-2017
		JP 2017510293 A	13-04-2017
		US 2017175136 A1	22-06-2017
		WO 2015138852 A1	17-09-2015

WO 2014043131 A1	20-03-2014	EP 2895611 A1	22-07-2015
		JP 2015529466 A	08-10-2015
		US 2015224209 A1	13-08-2015
		US 2017157270 A1	08-06-2017
		WO 2014043131 A1	20-03-2014

WO 2010046493 A2	29-04-2010	EP 2344641 A2	20-07-2011
		US 2011294873 A1	01-12-2011
		WO 2010046493 A2	29-04-2010

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 62045997 B [0001]
- US 4873192 A [0043]
- US 7541179 B [0102]
- US 5399346 A, Anderson [0142]
- US 20110301073 A [0153]
- US 2009054985 A [0160]
- US 5543158 A [0176]
- US 5641515 A [0176]
- US 5399363 A [0176]
- US 5756353 A [0184]
- US 5804212 A [0184]
- US 5725871 A [0184]
- US 5780045 A [0184]

Non-patent literature cited in the description

- SINGLETON et al. Dictionary of Microbiology and Molecular Biology. 1994 [0033]
- The Cambridge Dictionary of Science and Technology. 1988 [0033]
- The Glossary of Genetics. Springer Verlag, 1991 [0033]
- HALE ; MARHAM. *The Harper Collins Dictionary of Biology*, 1991 [0033]
- LI et al. *J. Biol. Chem.*, 1985, vol. 260, 14, , 901 [0035]
- LI et al. *Proc. Natl. Acad. Sci.*, 1990, vol. 87, 8207 [0035]
- SHEHEE et al. *J. Mol. Biol.*, 1989, vol. 205, 41 [0035]
- MARGOT et al. *J. Mol. Biol.*, 1989, vol. 205, 15 [0035]
- LI, Q. et al. *Genomics*, 1991, vol. 9, 488 [0035]
- WAHL, G. M. ; S. L. BERGER. *Methods Enzymol.*, 1987, vol. 152, 399 [0039]
- KIMMEL, A. R. *Methods Enzymol.*, 1987, vol. 152, 507 [0039]
- BENTON ; DAVIS. *Science*, 1977, vol. 196, 180 [0041]
- GRUNSTEIN ; ROGNESS. *Proc. Natl. Acad. Sci.*, 1975, vol. 72, 3961 [0041]
- AUSUBEL et al. *Current Protocols in Molecular Biology*. Wiley Interscience, 2001 [0041]
- BERGER ; KIMMEL. *Guide to Molecular Cloning Techniques*. Academic Press, 1987 [0041]
- SAMBROOK et al. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press [0041]
- KUNKEL. *Proc. Natl. Acad. Sci. USA.*, 1985, vol. 82, 488-492 [0043]
- KUNKEL et al. *Methods in Enzymol.*, 1987, vol. 154, 367-382 [0043]
- WATSON, J. D. et al. *Molecular Biology of the Gene*. Benjamin/Cummings, 1987 [0043]
- DAYHOFF et al. *Atlas of Protein Sequence and Structure*. Natl. Biomed. Res. Found, 1978 [0043]
- BURGESS-BEUSSE et al. *Proc. Natl. Acad. Sci. USA*, 2002, vol. 99, 16433 [0067]
- ZHAN et al. *Hum. Genet.*, 2001, vol. 109, 471 [0067]
- PRIOLEAU et al. *EMBO J.*, 1999, vol. 18, 4035-4048 [0068]
- CHUNG et al. *Cell*, 1993, vol. 74, 505-514 [0068]
- PIKAART et al. *Genes Dev.*, 1998, vol. 12, 2852-2862 [0068]
- WANG et al. *Nat. Biotechnol.*, 1997, vol. 15, 239-243 [0068]
- TABOIT-DAMERON et al. *Transgenic Res.*, 1999, vol. 8, 223-235 [0068]
- CHUNG et al. *Proc. Natl. Acad. Sci.*, 1997, vol. 94, 575-580 [0068]
- BELL et al. *Cell*, 1999, vol. 98, 387-396 [0068]
- BULGER et al. *PNAS*, 1999, vol. 96, 5129-5134 [0079]
- Molecular Mechanism of Beta Thalassemia. FORGET et al. *Disorders of Hemoglobin. Genetics, Pathophysiology and Clinical Management*. Cambridge University Press, 2001 [0079]
- Hemoglobin Switching. STAMATOYANNOPOULOS et al. *Molecular Basis of Blood Disorders*. W.B. Saunders, 2001 [0079]
- ANTONIOU et al. *EMBO J.*, 1988, vol. 7, 377-384 [0080]
- TRUDEL et al. *Genes Dev.*, 1987, vol. 1, 954-961 [0080]
- TRUDEL et al. *Mol. Cell. Biol.*, 1987, vol. 7, 4024-4029 [0080]
- GROSVELD et al. *Cell*, 1987, vol. 51, 975-985 [0080]
- TALBOT et al. *Nature*, 1989, vol. 338, 352 [0080]
- ELLIS et al. *EMBO J.*, 1996, vol. 15, 562-568 [0081]
- COLLIS et al. *EMBO J.*, 1990, vol. 9, 233-240 [0081]
- BUNGERT et al. *Mol. Cell Biol.*, 1999, vol. 19, 3062-3072 [0081]
- LEY et al. *Ann. N.Y. Acad. Sci.*, 1998, vol. 850, 45-53 [0081]
- HUG et al. *Mol. Cell Biol.*, 1996, vol. 26, 2906-2912 [0081]

- **CHANG et al.** *Stem cell reviews*, 2013, vol. 9, 397-407 [0097]
- **SADELAIN et al.** *Proc. Nat'l Acad. Sci.*, 1995, vol. 92, 6728-6732 [0102]
- **NAGEL et al.** *Proc. Natl. Acad. Sci. U.S.A.*, 1979, vol. 76, 670-672 [0120]
- **MILLER et al.** *Mol. Cell. Biol.*, 1985, vol. 5, 431-437 [0140]
- **MILLER et al.** *Mol. Cell. Biol.*, 1986, vol. 6, 2895-2902 [0140]
- **DANOS et al.** *Proc. Natl. Acad. Sci. USA*, 1988, vol. 85, 6460-6464 [0140]
- **BREGNI et al.** *Blood*, 1992, vol. 80, 1418-1422 [0141]
- **XU et al.** *Exp. Hemat.*, 1994, vol. 22, 223-230 [0141]
- **HUGHES et al.** *J. Clin. Invest.*, 1992, vol. 89, 1817 [0141]
- **CAYOUILLE et al.** *Human Gene Therapy*, 1997, vol. 8, 423-430 [0142]
- **KIDO et al.** *Current Eye Research*, 1996, vol. 15, 833-844 [0142]
- **BLOOMER et al.** *Journal of Virology*, 1997, vol. 71, 6641-6649 [0142]
- **NALDINI et al.** *Science*, 1996, vol. 272, 263-267 [0142]
- **MIYOSHI et al.** *Proc. Natl. Acad. Sci. U.S.A.*, 1997, vol. 94, 10319 [0142]
- **MILLER.** *Human Gene Therapy*, 1990, 15-14 [0142]
- **FRIEDMAN.** *Science*, 1989, vol. 244, 1275-1281 [0142]
- **EGLITIS et al.** *BioTechniques*, 1988, vol. 6, 608-614 [0142]
- **TOLSTOSHEV et al.** *Current Opinion in Biotechnology*, 1990, vol. 1, 55-61 [0142]
- **SHARP.** *The Lancet*, 1991, vol. 337, 1277-1278 [0142]
- **CORNETTA et al.** *Nucleic Acid Research and Molecular Biology*, 1987, vol. 36, 311-322 [0142]
- **ANDERSON.** *Science*, 1984, vol. 226, 401-409 [0142]
- **MOEN.** *Blood Cells*, 1991, vol. 17, 407-416 [0142]
- **MILLER et al.** *Biotechnology*, 1989, vol. 7, 980-990 [0142]
- **LE GAL LA SALLE et al.** *Science*, 1993, vol. 259, 988-990 [0142]
- **JOHNSON.** *Chest*, 1995, vol. 107, 77S-83S [0142]
- **ROSENBERG et al.** *N. Engl. J. Med*, 1990, vol. 323, 370 [0142]
- **KIM et al.** *Proc Nat'l Acad Sci USA*, 1996, vol. 93 (3), 1156-1160 [0153]
- **BELFORT et al.** *Nucleic Acids Res.*, 1997, vol. 25, 3379-3388 [0154]
- **Nucleases.** Cold Spring Harbor Laboratory Press, 1993 [0154]
- **JINEK et al.** *Science*, 2012, vol. 337, 816-821 [0157]
- **SADELAIN et al.** Safe Harbours for the integration of new DNA in the human genome. *Nature Reviews*, 2012, vol. 12, 51-58 [0158]
- **PAPAPETROU et al.** Genomic safe harbors permit high β -globin transgene expression in thalassemia induced pluripotent stem cells. *Nat Biotechnol.*, vol. 29 (1), 73-8 [0158]
- **ORY et al.** *Proc. Natl. Acad. Sci. USA*, 1996, vol. 93, 11382-11388 [0160]
- **DULL et al.** *J. Viral.*, 1998, vol. 72, 8463-8471 [0160]
- **ZUFFERY et al.** *J. Viral.*, 1998, vol. 72, 9873-9880 [0160]
- **FOLLENZI et al.** *Nature Genetics*, 2000, vol. 25, 217-222 [0160]
- **FEIGNER et al.** *Proc. Natl. Acad. Sci. U.S.A.*, 1987, vol. 84, 7413 [0161]
- **ONO et al.** *Neuroscience Letters*, 1990, vol. 17, 259 [0161]
- **BRIGHAM et al.** *Am. J. Med. Sci.*, 1989, vol. 298, 278 [0161]
- **STAUBINGER et al.** *Methods in Enzymology*, 1983, vol. 101, 512 [0161]
- **WU et al.** *Journal of Biological Chemistry*, 1988, vol. 263, 14621 [0161]
- **WU et al.** *Journal of Biological Chemistry*, 1989, vol. 264, 16985 [0161]
- **WOLFF et al.** *Science*, 1990, vol. 247, 1465 [0161]
- **MOI et al.** *Haematol*, 01 March 2008, vol. 93 (3), 325-330 [0169]
- **CHANG et al.** *Proc Natl Acad Sci USA*, 2006, vol. 103, 1036-40 [0169]
- **TAKAHASHI et al.** *Cell*, 2006, vol. 126, 663-76 [0169]
- **MEISSNER et al.** *Nat Biotechnol*, 2007, vol. 25, 1177-81 [0169]
- **NAKAGAWA et al.** *Nat Biotechnol*, 2007, vol. 26, 101-6 [0169]
- **OKITA et al.** *Nature*, 2007, vol. 448, 313-7 [0169]
- **PARK et al.** *Nature*, 2007, vol. 451, 141-6 [0169]
- **TAKAHASHI et al.** *Nat Protoc*, 2007, vol. 2, 3081-9 [0169]
- **TAKAHASHI K et al.** *Cell*, 2007, vol. 131, 861-72 [0169]
- **WERNIG et al.** *Nature*, 2007, vol. 448, 318-24 [0169]
- **YU J et al.** *Science*, 2007, vol. 318, 1917-20 [0169]
- **HANNA et al.** *Science*, 2007, vol. 318, 1920-3 [0169]
- **REMINGTON'S PHARMACEUTICAL SCIENCE.** 1985 [0180]
- **SAMBROOK.** *Molecular Cloning: A Laboratory Manual.* 1989 [0207]
- **GAIT.** *Oligonucleotide Synthesis*, 1984 [0207]
- **FRESHNEY.** *Animal Cell Culture*, 1987 [0207]
- **Methods in Enzymology.** **WEIR.** *Handbook of Experimental Immunology.* 1996 [0207]
- **MILLER ; CALOS.** *Gene Transfer Vectors for Mammalian Cells*, 1987 [0207]
- **AUSUBEL.** *Current Protocols in Molecular Biology*, 1987 [0207]
- **MULLIS.** *PCR: The Polymerase Chain Reaction*, 1994 [0207]

- **COLIGAN.** *Current Protocols in Immunology*, 1991 [0207]
- **WEATHERALL, D.J. ; CLEGG, J.B.** The Thalassemia Syndrome. Blackwell Scientific, 1981 [0222]
- **STAMATOYANNOPOULOS, G. ; NIENHUIS, A.W. ; MAJERUS, P. ; VARMUS, H.** The Molecular Basis of Blood Diseases. WB Saunders, 1994 [0222]
- **WEATHERALL, D.J.** Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. *Nat Rev Genet*, 2001, vol. 2, 245-255 [0222]
- **STEINBERG, M.H. ; FORGET, B.G. ; HIGGS, D.R. ; NAGEL, R.L.** Molecular Mechanism of β Thalassemia. Cambridge University Press, 2001 [0222]
- **COOLEY, T.B. ; LEE, P. A.** series of cases of splenomegaly in children with anemia and peculiar bone changes. *Trans. Am. Pediatr. Soc.*, 1925, vol. 37, 29 [0222]
- **GIARDINA, P.J. ; GRADY, R.W.** Chelation therapy in beta-thalassemia: an optimistic update. *Semin Hematol*, 2001, vol. 38, 360-366 [0222]
- **GIARDINI, C. ; LUCARELLI, G.** Bone marrow transplantation in the treatment of thalassemia. *Current opinion in hematology*, 1994, vol. 1, 170-176 [0222]
- **BOULAD, F. ; GIARDINA, P. ; GILLIO, A. ; KERNAN, N. ; SMALL, T. ; BROCHSTEIN, J. ; VAN SYCKLE, K. ; GEORGE, D. ; SZABOLCS, P. ; O'REILLY, R.J.** Bone marrow transplantation for homozygous beta-thalassemia. The Memorial Sloan-Kettering Cancer Center experience. *Ann N Y Acad Sci*, 1998, vol. 850, 498-502 [0222]
- **LUCARELLI, G. ; CLIFT, R.A. ; ALIMBERTI, M. ; ANGELUCCI, E. ; GIARDINI, C. ; BARONCIANI, D. ; POLCHI, P. ; ANDREANI, M. ; GAZIEV, D. ; ERER, B.** Bone marrow transplantation in adult thalassemic patients. *Blood*, 1999, vol. 93, 1164-1167 [0222]
- **TISDALE, J. ; SADELAIN, M.** Toward gene therapy for disorders of globin synthesis. *Semin Hematol*, 2001, vol. 38, 382-392 [0222]
- **PAULING, L. ; ITANO, H.A. ; SINGER, S.J. ; WELLS, I.C.** Sickle cell anemia, a molecular disease. *Science*, 1949, vol. 110, 543-546 [0222]
- **SWANK, R.A. ; STAMATOYANNOPOULOS, G.** Fetal gene reactivation. *Curr Opin Genet Dev*, 1998, vol. 8, 366-370 [0222]
- **PLATT, O.S. ; ORKIN, S.H. ; DOVER, G. ; BEARDSLEY, G.P. ; MILLER, B. ; NATHAN, D.G.** Hydroxyurea enhances fetal hemoglobin production in sickle cell anemia. *J Clin Invest*, 1984, vol. 74, 652-656 [0222]
- **CHARACHE, S. ; DOVER, G.J. ; MOORE, R.D. ; ECKERT, S. ; BALLAS, S.K. ; KOSHY, M. ; MILLNER, P.F. ; ORRINGER, E.P. ; PHILLIPS, G., JR. ; PLATT, O.S. et al.** Hydroxyurea: effects on hemoglobin F production in patients with sickle cell anemia. *Blood*, 1992, vol. 79, 2555-2565 [0222]
- **ATWEH, G.F. ; LOUKOPOULOS, D.** Pharmacological induction of fetal hemoglobin in sickle cell disease and beta-thalassemia. *Semin Hematol*, 2001, vol. 38, 367-373 [0222]
- **PERRINE, S.P. ; CASTANEDA, S.A. ; BOOSALIS, M.S. ; WHITE, G.L. ; JONES, B.M. ; BOHACEK, R.** Induction of fetal globin in beta-thalassemia: Cellular obstacles and molecular progress. *Ann N Y Acad Sci*, 2005, vol. 1054, 257-265 [0222]
- **STAMATOYANNOPOULOS, G.** Prospects for developing a molecular cure for thalassemia. *Hematology*, 2005, vol. 10 (1), 255-257 [0222]
- **VERMYLEN, C. ; CORNU, G. ; FERSTER, A. ; BRICHARD, B. ; NINANE, J. ; FERRANT, A. ; ZENEBERGH, A. ; MAES, P. ; DHOOGHE, C. ; BENOIT, Y.** Haematopoietic stem cell transplantation for sickle cell anaemia: the first 50 patients transplanted in Belgium. *Bone Marrow Transplant*, 1998, vol. 22, 1-6 [0222]
- **LUZZATTO, L. ; GOODFELLOW, P.** Sickle cell anaemia. A simple disease with no cure. *Nature*, 1989, vol. 337, 17-18 [0222]
- **SADELAIN, M.** Genetic treatment of the haemoglobinopathies: recombinations and new combinations. *Br J Haematol*, 1997, vol. 98, 247-253 [0222]
- **SADELAIN, M. ; BOULAD, F. ; GALANELLO, R. ; GIARDINA, P. ; LOCATELLI, F. ; MAGGIO, A. ; RIVELLA, S. ; RIVIERE, I. ; TISDALE, J.** Therapeutic options for patients with severe beta-thalassemia: the need for globin gene therapy. *Hum Gene Ther*, 2007, vol. 18, 1-9 [0222]
- **BORGNA-PIGNATTI, C. ; RUGOLOTTI, S. ; DE STEFANO, P. ; ZHAO, H. ; CAPPELLINI, M.D. ; DEL VECCHIO, G.C. ; ROMEO, M.A. ; FORNI, G.L. ; GAMBERINI, M.R. ; GHILARDI, R.** Survival and complications in patients with thalassemia major treated with transfusion and deferoxamine. *Haematologica*, 2004, vol. 89, 1187-1193 [0222]
- **TELFER, P.T. ; WARBURTON, F. ; CHRISTOU, S. ; HADJIGAVRIEL, M. ; SITAROU, M. ; KOLNAGOU, A. ; ANGASTINIOTIS, M.** Improved survival in thalassemia major patients on switching from desferrioxamine to combined chelation therapy with desferrioxamine and deferiprone. *Haematologica*, 2009, vol. 94, 1777-1778 [0222]
- **LADIS, V. ; CHOULIARAS, G. ; BERDOUKAS, V. ; CHATZILIAMI, A. ; FRAGODIMITRI, C. ; KARABATSOS, F. ; YOUSSEF, J. ; KATTAMIS, A. ; KARAGIORGA-LAGANA, M.** Survival in a large cohort of Greek patients with transfusion-dependent beta thalassaemia and mortality ratios compared to the general population. *European journal of haematology*, 2011, vol. 86, 332-338 [0222]
- **MANCUSO, A. ; SCIARRINO, E. ; RENDA, M.C. ; MAGGIO, A.** A prospective study of hepatocellular carcinoma incidence in thalassemia. *Hemoglobin*, 2006, vol. 30, 119-124 [0222]

- **PERSONS, D.A. ; TISDALE, J.F.** Gene therapy for the hemoglobin disorders. *Semin Hematol*, 2004, vol. 41, 279-286 [0222]
- **SADELAIN, M.** Recent advances in globin gene transfer for the treatment of beta-thalassemia and sickle cell anemia. *Current opinion in hematology*, 2006, vol. 13, 142-148 [0222]
- **MAY, C. ; RIVELLA, S. ; CALLEGARI, J. ; HELLER, G. ; GAENSLER, K.M. ; LUZZATTO, L. ; SADELAIN, M.** Therapeutic haemoglobin synthesis in beta-thalassaemic mice expressing lentivirus-encoded human beta-globin. *Nature*, 2000, vol. 406, 82-86 [0222]
- **MAY, C. ; RIVELLA, S. ; CHADBURN, A. ; SADELAIN, M.** Successful treatment of murine beta-thalassemia intermedia by transfer of the human beta-globin gene. *Blood*, 2002, vol. 99, 1902-1908 [0222]
- **RIVELLA, S. ; MAY, C. ; CHADBURN, A. ; RIVIERE, I. ; SADELAIN, M.** A novel murine model of Cooley anemia and its rescue by lentiviral-mediated human beta-globin gene transfer. *Blood*, 2003, vol. 101, 2932-2939 [0222]
- **SADELAIN, M. ; BOULAD, F. ; LISOWSKI, L. ; MOI, P. ; RIVIERE, I.** Stem cell engineering for the treatment of severe hemoglobinopathies. *Curr Mol Med*, 2008, vol. 8, 690-697 [0222]
- **BANK, A. ; DORAZIO, R. ; LEBOULCH, P.** A phase I/II clinical trial of beta-globin gene therapy for beta-thalassemia. *Ann N Y Acad Sci*, 2005, vol. 1054, 308-316 [0222]
- **CAVAZZANA-CALVO, M. ; PAYEN, E. ; NEGRE, O. ; WANG, G. ; HEHIR, K. ; FUSIL, F. ; DOWN, J. ; DENARO, M. ; BRADY, T. ; WESTERMAN, K.** Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia. *Nature*, 2010, vol. 467, 318-322 [0222]
- **BRAUN, C.J. ; BOZTUG, K. ; PARUZYSKI, A. ; WITZEL, M. ; SCHWARZER, A. ; ROTHE, M. ; MODLICH, U. ; BEIER, R. ; GOHRING, G. ; STEINEMANN, D.** Gene therapy for Wiskott-Aldrich syndrome—long-term efficacy and genotoxicity. *Sci Transl Med*, 2014, vol. 6, 227ra233 [0222]
- **CHANG, A.H. ; SADELAIN, M.** The genetic engineering of hematopoietic stem cells: the rise of lentiviral vectors, the conundrum of the Itr, and the promise of lineage-restricted vectors. *Mol Ther*, 2007, vol. 15, 445-456 [0222]
- **PAWLIUK, R. ; WESTERMAN, K.A. ; FABRY, M.E. ; PAYEN, E. ; TIGHE, R. ; BOUHASSIRA, E.E. ; ACHARYA, S.A. ; ELLIS, J. ; LONDON, I.M. ; EAVES, C.J.** Correction of sickle cell disease in transgenic mouse models by gene therapy. *Science*, 2001, vol. 294, 2368-2371 [0222]
- **EMERY, D.W. ; CHEN, H. ; LI, Q. ; STAMATOYANNOPOULOS, G.** Development of a condensed locus control region cassette and testing in retrovirus vectors for A gamma-globin. *Blood Cells Mol Dis*, 1998, vol. 24, 322-339 [0222]
- **MICCIO, A. ; CESARI, R. ; LOTTI, F. ; ROSSI, C. ; SANVITO, F. ; PONZONI, M. ; ROUTLEDGE, S.J. ; CHOW, C.M. ; ANTONIOU, M.N. ; FERRARI, G.** In vivo selection of genetically modified erythroblastic progenitors leads to long-term correction of beta-thalassemia. *Proc Natl Acad Sci USA*, 2008, vol. 105, 10547-10552 [0222]
- **SADELAIN, M. ; WANG, C.H. ; ANTONIOU, M. ; GROSVELD, F. ; MULLIGAN, R.C.** Generation of a high-titer retroviral vector capable of expressing high levels of the human beta-globin gene. *Proc Natl Acad Sci USA*, 1995, vol. 92, 6728-6732 [0222]
- **SAMAKOGLU, S. ; LISOWSKI, L. ; BUDAK-ALPDOGAN, T. ; USACHENKO, Y. ; ACUTO, S. ; DI MARZO, R. ; MAGGIO, A. ; ZHU, P. ; TISDALE, J.F. ; RIVIERE, I.** A genetic strategy to treat sickle cell anemia by coregulating globin transgene expression and RNA interference. *Nat Biotechnol*, 2006, vol. 24, 89-94 [0222]
- **PESTINA, T.I. ; HARGROVE, P.W. ; JAY, D. ; GRAY, J.T. ; BOYD, K.M. ; PERSONS, D.A.** Correction of murine sickle cell disease using gamma-globin lentiviral vectors to mediate high-level expression of fetal hemoglobin. *Mol Ther*, 2009, vol. 17, 245-252 [0222]
- **HANAWA, H. ; YAMAMOTO, M. ; ZHAO, H. ; SHIMADA, T. ; PERSONS, D.A.** Optimized lentiviral vector design improves titer and transgene expression of vectors containing the chicken beta-globin locus HS4 insulator element. *Mol Ther*, 2009, vol. 17, 667-674 [0222]
- **ARUMUGAM, P.I. ; SCHOLLES, J. ; PERELMAN, N. ; XIA, P. ; YEE, J.K. ; MALIK, P.** Improved human beta-globin expression from self-inactivating lentiviral vectors carrying the chicken hypersensitive site-4 (cHS4) insulator element. *Mol Ther*, 2007, vol. 15, 1863-1871 [0222]
- **FRASER, P. ; PRUZINA, S. ; ANTONIOU, M. ; GROSVELD, F.** Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes. *Genes & development*, 1993, vol. 7, 106-113 [0222]
- **NAVAS, P.A. ; PETERSON, K.R. ; LI, Q. ; SKARPIDI, E. ; ROHDE, A. ; SHAW, S.E. ; CLEGG, C.H. ; ASANO, H. ; STAMATOYANNOPOULOS, G.** Developmental specificity of the interaction between the locus control region and embryonic or fetal globin genes in transgenic mice with an HS3 core deletion. *Molecular and cellular biology*, 1998, vol. 18, 4188-4196 [0222]

- **LI, Q. ; STAMATOYANNOPOULOS, G.** Hypersensitive site 5 of the human beta locus control region functions as a chromatin insulator. *Blood*, 1994, vol. 84, 1399-1401 [0222]
- **LI, Q. ; ZHANG, M. ; HAN, H. ; ROHDE, A. ; STAMATOYANNOPOULOS, G.** Evidence that DNase I hypersensitive site 5 of the human beta-globin locus control region functions as a chromosomal insulator in transgenic mice. *Nucleic Acids Res*, 2002, vol. 30, 2484-2491 [0222]
- **PUTHENVEETIL, G. ; SCHOLLES, J. ; CARBONELL, D. ; QURESHI, N. ; XIA, P. ; ZENG, L. ; LI, S. ; YU, Y. ; HITI, A.L. ; YEE, J.K.** Successful correction of the human beta-thalassemia major phenotype using a lentiviral vector. *Blood*, 2004, vol. 104, 3445-3453 [0222]
- **WILBER, A. ; NIENHUIS, A.W. ; PERSONS, D.A.** Transcriptional regulation of fetal to adult hemoglobin switching: new therapeutic opportunities. *Blood*, 2011, vol. 117, 3945-3953 [0222]
- **ARUMUGAM, P.I. ; HIGASHIMOTO, T. ; URBINATI, F. ; MODLICH, U. ; NESTHEIDE, S. ; XIA, P. ; FOX, C. ; CORSINOTTI, A. ; BAUM, C. ; MALIK, P.** Genotoxic potential of lineage-specific lentivirus vectors carrying the beta-globin locus control region. *Mol Ther*, 2009, vol. 17, 1929-1937 [0222]
- **CHANG, K.H. ; FANG, X. ; WANG, H. ; HUANG, A. ; CAO, H. ; YANG, Y. ; BONIG, H. ; STAMATOYANNOPOULOS, J.A. ; PAPAYANNOPOULOU, T.** Epigenetic modifications and chromosome conformations of the beta globin locus throughout development. *Stem cell reviews*, 2013, vol. 9, 397-407 [0222]
- **PAPAYANNOPOULOU, T. ; PRIESTLEY, G.V. ; ROHDE, A. ; PETERSON, K.R. ; NAKAMOTO, B.** Hemopoietic lineage commitment decisions: in vivo evidence from a transgenic mouse model harboring micro LCR-beta_{pro}-LacZ as a transgene. *Blood*, 2000, vol. 95, 1274-1282 [0222]
- **NIENHUIS, A.W.** Development of gene therapy for blood disorders: an update. *Blood*, 2013, vol. 122, 1556-1564 [0222]
- **BAUM, C. ; KUSTIKOVA, O. ; MODLICH, U. ; LI, Z. ; FEHSE, B.** Mutagenesis and oncogenesis by chromosomal insertion of gene transfer vectors. *Hum Gene Ther*, 2006, vol. 17, 253-263 [0222]
- **NIENHUIS, A.W. ; DUNBAR, C.E. ; SORRENTINO, B.P.** Genotoxicity of retroviral integration in hematopoietic cells. *Mol Ther*, 2006, vol. 13, 1031-1049 [0222]
- **EMERY, D.W.** The use of chromatin insulators to improve the expression and safety of integrating gene transfer vectors. *Hum Gene Ther*, 2011, vol. 22, 761-774 [0222]
- **EVANS-GALEA, M.V. ; WIELGOSZ, M.M. ; HANAWA, H. ; SRIVASTAVA, D.K. ; NIENHUIS, A.W.** Suppression of clonal dominance in cultured human lymphoid cells by addition of the cHS4 insulator to a lentiviral vector. *Mol Ther*, 2007, vol. 15, 801-809 [0222]
- **RIVELLA, S. ; CALLEGARI, J.A. ; MAY, C. ; TAN, C.W. ; SADELAIN, M.** The cHS4 insulator increases the probability of retroviral expression at random chromosomal integration sites. *J Virol*, 2000, vol. 74, 4679-4687 [0222]
- **EMERY, D.W. ; YANNAKI, E. ; TUBB, J. ; STAMATOYANNOPOULOS, G.** A chromatin insulator protects retrovirus vectors from chromosomal position effects. *Proc Natl Acad Sci USA*, 2000, vol. 97, 9150-9155 [0222]
- **EMERY, D.W. ; YANNAKI, E. ; TUBB, J. ; NISHINO, T. ; LI, Q. ; STAMATOYANNOPOULOS, G.** Development of virus vectors for gene therapy of beta chain hemoglobinopathies: flanking with a chromatin insulator reduces gamma-globin gene silencing in vivo. *Blood*, 2002, vol. 100, 2012-2019 [0222]
- **YANNAKI, E. ; TUBB, J. ; AKER, M. ; STAMATOYANNOPOULOS, G. ; EMERY, D.W.** Topological constraints governing the use of the chicken HS4 chromatin insulator in oncoretrovirus vectors. *Mol Ther*, 2002, vol. 5, 589-598 [0222]
- **HINO, S. ; FAN, J. ; TAGUWA, S. ; AKASAKA, K. ; MATSUOKA, M.** Sea urchin insulator protects lentiviral vector from silencing by maintaining active chromatin structure. *Gene Ther*, 2004, vol. 11, 819-828 [0222]
- **RAMEZANI, A. ; HAWLEY, T.S. ; HAWLEY, R.G.** Performance- and safety-enhanced lentiviral vectors containing the human interferon-beta scaffold attachment region and the chicken beta-globin insulator. *Blood*, 2003, vol. 101, 4717-4724 [0222]
- **RAMEZANI, A. ; HAWLEY, T.S. ; HAWLEY, R.G.** Combinatorial incorporation of enhancer-blocking components of the chicken beta-globin 5'HS4 and human T-cell receptor alpha/delta BEAD-1 insulators in self-inactivating retroviral vectors reduces their genotoxic potential. *Stem Cells*, 2008, vol. 26, 3257-3266 [0222]
- **YANNAKI, E. ; EMERY, D.W. ; STAMATOYANNOPOULOS, G.** Gene therapy for beta-thalassemia: the continuing challenge. *Expert reviews in molecular medicine*, 2010, vol. 12, e31 [0222]
- **PERSONS, D.A.** The challenge of obtaining therapeutic levels of genetically modified hematopoietic stem cells in beta-thalassemia patients. *Ann N Y Acad Sci*, 2010, vol. 1202, 69-74 [0222]
- **PERUMBETI, A. ; MALIK, P.** Therapy for beta-globinopathies: a brief review and determinants for successful and safe correction. *Ann N Y Acad Sci*, 2010, vol. 1202, 36-44 [0222]

- **JOHNSON, K.D. ; GRASS, J.A. ; PARK, C. ; IM, H. ; CHOI, K. ; BRESNICK, E.H.** Highly restricted localization of RNA polymerase II within a locus control region of a tissue-specific chromatin domain. *Molecular and cellular biology*, 2003, vol. 23, 6484-6493 [0222]
- **VIEIRA, K.F. ; LEVINGS, P.P. ; HILL, M.A. ; CRUSSELLE, V.J. ; KANG, S.H. ; ENGEL, J.D. ; BUNGERT, J.** Recruitment of transcription complexes to the beta-globin gene locus in vivo and in vitro. *J Biol Chem*, 2004, vol. 279, 50350-50357 [0222]
- **LEVINGS, P.P. ; ZHOU, Z. ; VIEIRA, K.F. ; CRUSSELLE-DAVIS, V.J. ; BUNGERT, J.** Recruitment of transcription complexes to the beta-globin locus control region and transcription of hypersensitive site 3 prior to erythroid differentiation of murine embryonic stem cells. *The FEBS journal*, 2006, vol. 273, 746-755 [0222]
- **FELSENFELD, G. ; GROUDINE, M.** Controlling the double helix. *Nature*, 2003, vol. 421, 448-453 [0222]
- **FELSENFELD, G.** Chromatin as an essential part of the transcriptional mechanism. *Nature*, 1992, vol. 355, 219-224 [0222]
- **BROWNELL, J.E. ; ALLIS, C.D.** Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. *Curr Opin Genet Dev*, 1996, vol. 6, 176-184 [0222]
- **KINGSTON, R.E. ; NARLIKAR, G.J.** ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. *Genes & development*, 1999, vol. 13, 2339-2352 [0222]
- **TSUKIYAMA, T. ; WU, C.** Chromatin remodeling and transcription. *Curr Opin Genet Dev*, 1997, vol. 7, 182-191 [0222]
- **WOLFFE, A.P. ; WONG, J. ; PRUSS, D.** Activators and repressors: making use of chromatin to regulate transcription. *Genes to cells : devoted to molecular & cellular mechanisms*, 1997, vol. 2, 291-302 [0222]
- **KADONAGA, J.T.** Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. *Cell*, 1998, vol. 92, 307-313 [0222]
- **STRUHL, K.** Histone acetylation and transcriptional regulatory mechanisms. *Genes & development*, 1998, vol. 12, 599-606 [0222]
- **GROSS, D.S. ; GARRARD, W.T.** Nuclease hypersensitive sites in chromatin. *Annual review of biochemistry*, 1988, vol. 57, 159-197 [0222]
- **ELGIN, S.C.** Anatomy of hypersensitive sites. *Nature*, 1984, vol. 309, 213-214 [0222]
- **WU, C.** The 5' ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. *Nature*, 1980, vol. 286, 854-860 [0222]
- **FELSENFELD, G. ; BOYES, J. ; CHUNG, J. ; CLARK, D. ; STUDITSKY, V.** Chromatin structure and gene expression. *Proc Natl Acad Sci U S A*, 1996, vol. 93, 9384-9388 [0222]
- **BURGESS-BEUSSE, B. ; FARRELL, C. ; GASZNER, M. ; LITT, M. ; MUTSKOV, V. ; RECILLAS-TARGA, F. ; SIMPSON, M. ; WEST, A. ; FELSENFELD, G.** The insulation of genes from external enhancers and silencing chromatin. *Proc Natl Acad Sci U S A*, 2002, vol. 99 (4), 16433-16437 [0222]
- **ELGIN, S.C.** DNAase I-hypersensitive sites of chromatin. *Cell*, 1981, vol. 27, 413-415 [0222]
- **MCGHEE, J.D. ; WOOD, W.I. ; DOLAN, M. ; ENGEL, J.D. ; FELSENFELD, G.** A 200 base pair region at the 5' end of the chicken adult beta-globin gene is accessible to nuclease digestion. *Cell*, 1981, vol. 27, 45-55 [0222]
- **LOWREY, C.H. ; BODINE, D.M. ; NIENHUIS, A.W.** Mechanism of DNase I hypersensitive site formation within the human globin locus control region. *Proc Natl Acad Sci USA*, 1992, vol. 89, 1143-1147 [0222]
- **ADAMS, C.C. ; WORKMAN, J.L.** Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Molecular and cellular biology*, 1995, vol. 15, 1405-1421 [0222]
- **MCARTHUR, M. ; GERUM, S. ; STAMATOYANOPOULOS, G.** Quantification of DNaseI-sensitivity by real-time PCR: quantitative analysis of DNaseI-hypersensitivity of the mouse beta-globin LCR. *J Mol Biol*, 2001, vol. 313, 27-34 [0222]
- **DORSCHNER, M.O. ; HAWRYLYCZ, M. ; HUMBERT, R. ; WALLACE, J.C. ; SHAFER, A. ; KAWAMOTO, J. ; MACK, J. ; HALL, R. ; GOLDY, J. ; SABO, P.J.** High-throughput localization of functional elements by quantitative chromatin profiling. *Nat Methods*, 2004, vol. 1, 219-225 [0222]
- **SABO, P.J. ; KUEHN, M.S. ; THURMAN, R. ; JOHNSON, B.E. ; JOHNSON, E.M. ; CAO, H. ; YU, M. ; ROSENZWEIG, E. ; GOLDY, J. ; HAYDOCK, A.** Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat Methods*, 2006, vol. 3, 511-518 [0222]
- **SABO, P.J. ; HAWRYLYCZ, M. ; WALLACE, J.C. ; HUMBERT, R. ; YU, M. ; SHAFER, A. ; KAWAMOTO, J. ; HALL, R. ; MACK, J. ; DORSCHNER, M.O.** Discovery of functional noncoding elements by digital analysis of chromatin structure. *Proc Natl Acad Sci USA*, 2004, vol. 101, 16837-16842 [0222]
- **SABO, P.J. ; HUMBERT, R. ; HAWRYLYCZ, M. ; WALLACE, J.C. ; DORSCHNER, M.O. ; MCARTHUR, M. ; STAMATOYANOPOULOS, J.A.** Genome-wide identification of DNaseI hypersensitive sites using active chromatin sequence libraries. *Proc Natl Acad Sci USA*, 2004, vol. 101, 4537-4542 [0222]
- **THURMAN, R.E. ; RYNES, E. ; HUMBERT, R. ; VIERSTRA, J. ; MAURANO, M.T. ; HAUGEN, E. ; SHEFFIELD, N.C. ; STERGACHIS, A.B. ; WANG, H. ; VERNOT, B.** The accessible chromatin landscape of the human genome. *Nature*, 2012, vol. 489, 75-82 [0222]

- **STERGACHIS, A.B. ; NEPH, S. ; REYNOLDS, A. ; HUMBERT, R. ; MILLER, B. ; PAIGE, S.L. ; VERNOT, B. ; CHENG, J.B. ; THURMAN, R.E. ; SANDSTROM, R.** Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. *Cell*, 2013, vol. 154, 888-903 [0222]
- **NEPH, S. ; STERGACHIS, A.B. ; REYNOLDS, A. ; SANDSTROM, R. ; BORENSTEIN, E. ; STAMATOYANNOPOULOS, J.A.** Circuitry and dynamics of human transcription factor regulatory networks. *Cell*, 2012, vol. 150, 1274-1286 [0222]
- **MAURANO, M.T. ; HUMBERT, R. ; RYNES, E. ; THURMAN, R.E. ; HAUGEN, E. ; WANG, H. ; REYNOLDS, A.P. ; SANDSTROM, R. ; QU, H. ; BRODY, J.** Systematic localization of common disease-associated variation in regulatory DNA. *Science*, 2012, vol. 337, 1190-1195 [0222]
- **STERGACHIS, A.B. ; HAUGEN, E. ; SHAFER, A. ; FU, W. ; VERNOT, B. ; REYNOLDS, A. ; RAUBITSCHKE, A. ; ZIEGLER, S. ; LEPROUST, E.M. ; AKEY, J.M.** Exonic transcription factor binding directs codon choice and affects protein evolution. *Science*, 2013, vol. 342, 1367-1372 [0222]
- **NEPH, S. ; VIERSTRA, J. ; STERGACHIS, A.B. ; REYNOLDS, A.P. ; HAUGEN, E. ; VERNOT, B. ; THURMAN, R.E. ; JOHN, S. ; SANDSTROM, R. ; JOHNSON, A.K.** An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature*, 2012, vol. 489, 83-90 [0222]
- **RAMEZANI, A. ; HAWLEY, T.S. ; HAWLEY, R.G.** Stable gammaretroviral vector expression during embryonic stem cell-derived in vitro hematopoietic development. *Mol Ther*, 2006, vol. 14, 245-254 [0222]
- **RECILLAS-TARGA, F. ; PIKAART, M.J. ; BURGESS-BEUSSE, B. ; BELL, A.C. ; LITT, M.D. ; WEST, A.G. ; GASZNER, M. ; FELSENFELD, G.** Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc Natl Acad Sci USA*, 2002, vol. 99, 6883-6888 [0222]
- **GASZNER, M. ; FELSENFELD, G.** Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet*, 2006, vol. 7, 703-713 [0222]
- **WALLACE, J.A. ; FELSENFELD, G.** We gather together: insulators and genome organization. *Curr Opin Genet Dev*, 2007, vol. 17, 400-407 [0222]
- **CHUNG, J.H. ; BELL, A.C. ; FELSENFELD, G.** Characterization of the chicken beta-globin insulator. *Proc Natl Acad Sci U S A*, 1997, vol. 94, 575-580 [0222]
- **BELL, A.C. ; WEST, A.G. ; FELSENFELD, G.** The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*, 1999, vol. 98, 387-396 [0222]
- **RYU, B.Y. ; PERSONS, D.A. ; EVANS-GALEA, M.V. ; GRAY, J.T. ; NIENHUIS, A.W.** A chromatin insulator blocks interactions between globin regulatory elements and cellular promoters in erythroid cells. *Blood Cells MolDis*, 2007, vol. 39, 221-228 [0222]
- **RYU, B.Y. ; EVANS-GALEA, M.V. ; GRAY, J.T. ; BODINE, D.M. ; PERSONS, D.A. ; NIENHUIS, A.W.** An experimental system for the evaluation of retroviral vector design to diminish the risk for proto-oncogene activation. *Blood*, 2008, vol. 111, 1866-1875 [0222]
- **YAO, S. ; OSBORNE, C.S. ; BHARADWAJ, R.R. ; PASCERI, P. ; SUKONNIK, T. ; PANNELL, D. ; RECILLAS-TARGA, F. ; WEST, A.G. ; ELLIS, J.** Retrovirus silencer blocking by the cHS4 insulator is CTCF independent. *Nucleic Acids Res*, 2003, vol. 31, 5317-5323 [0222]
- **NISHINO, T. ; TUBB, J. ; EMERY, D.W.** Partial correction of murine beta-thalassemia with a gammaretrovirus vector for human gamma-globin. *Blood Cells MolDis*, 2006, vol. 37, 1-7 [0222]
- **AKER, M. ; TUBB, J. ; GROTH, A.C. ; BUKOVSKY, A.A. ; BELL, A.C. ; FELSENFELD, G. ; KIEM, H.P. ; STAMATOYANNOPOULOS, G. ; EMERY, D.W.** Extended core sequences from the cHS4 insulator are necessary for protecting retroviral vectors from silencing position effects. *Hum Gene Ther*, 2007, vol. 18, 333-343 [0222]
- **LI, C.L. ; EMERY, D.W.** The cHS4 chromatin insulator reduces gammaretroviral vector silencing by epigenetic modifications of integrated provirus. *Gene Ther*, 2008, vol. 15, 49-53 [0222]
- **MA, Y. ; RAMEZANI, A. ; LEWIS, R. ; HAWLEY, R.G. ; THOMSON, J.A.** High-level sustained transgene expression in human embryonic stem cells using lentiviral vectors. *Stem Cells*, 2003, vol. 21, 111-117 [0222]
- **CHANG, L.J. ; LIU, X. ; HE, J.** Lentiviral siRNAs targeting multiple highly conserved RNA sequences of human immunodeficiency virus type 1. *Gene Ther*, 2005, vol. 12, 1133-1144 [0222]
- **PLUTA, K. ; LUCE, M.J. ; BAO, L. ; AGHA-MOHAMMADI, S. ; REISER, J.** Tight control of transgene expression by lentivirus vectors containing second-generation tetracycline-responsive promoters. *J Gene Med*, 2005, vol. 7, 803-817 [0222]
- **JAKOBSSON, J. ; ROSENQVIST, N. ; THOMPSON, L. ; BARRAUD, P. ; LUNDBERG, C.** Dynamics of transgene expression in a neural stem cell line transduced with lentiviral vectors incorporating the cHS4 insulator. *Experimental cell research*, 2004, vol. 298, 611-623 [0222]

- **LEBOULCH, P. ; HUANG, G.M. ; HUMPHRIES, R.K. ; OH, Y.H. ; EAVES, C.J. ; TUAN, D.Y. ; LONDON, I.M.** Mutagenesis of retroviral vectors transducing human beta-globin gene and beta-globin locus control region derivatives results in stable transmission of an active transcriptional structure. *EMBO J*, 1994, vol. 13, 3065-3076 [0222]
- **KIM, T.H. ; ABDULLAEV, Z.K. ; SMITH, A.D. ; CHING, K.A. ; LOUKINOV, D.I. ; GREEN, R.D. ; ZHANG, M.Q. ; LOBANENKOV, V.V. ; REN, B.** Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell*, 2007, vol. 128, 1231-1245 [0222]
- **YUSUFZAI, T.M. ; FELSENFELD, G.** The 5'-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element. *Proc Natl Acad Sci U S A*, 2004, vol. 101, 8620-8624 [0222]
- **PHILLIPS, J.E. ; CORCES, V.G.** CTCF: master weaver of the genome. *Cell*, 2009, vol. 137, 1194-1211 [0222]
- **GILES, K.E. ; GOWHER, H. ; GHIRLANDO, R. ; JIN, C. ; FELSENFELD, G.** Chromatin boundaries, insulators, and long-range interactions in the nucleus. *Cold Spring Harbor symposia on quantitative biology*, 2010, vol. 75, 79-85 [0222]
- **BARSKI, A. ; CUDDAPAH, S. ; CUI, K. ; ROH, T.Y. ; SCHONES, D.E. ; WEI, G. ; CHEPELEV, I. ; ZHAO, K.** High-resolution profiling of histone methylations in the human genome. *Cell*, 2007, vol. 129, 823-837 [0222]
- **WANG, H. ; MAURANO, M.T. ; QU, H. ; VARLEY, K.E. ; GERTZ, J. ; PAULI, F. ; LEE, K. ; CANFIELD, T. ; WEAVER, M. ; SANDSTROM, R.** Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome research*, 2012, vol. 22, 1680-1688 [0222]
- **SCHMIDT, D. ; SCHWALIE, P.C. ; WILSON, M.D. ; BALLESTER, B. ; GONCALVES, A. ; KUTTER, C. ; BROWN, G.D. ; MARSHALL, A. ; FLICEK, P. ; ODOM, D.T.** Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell*, 2012, vol. 148, 335-348 [0222]
- **REDA, M. ; BAGLIVO, I. ; BURGESS-BEUSSE, B. ; ESPOSITO, S. ; FATTORUSSO, R. ; FELSENFELD, G. ; PEDONE, P.V.** Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. *J Biol Chem*, 2007, vol. 282, 33336-33345 [0222]
- **DICKSON, J. ; GOWHER, H. ; STROGANTSEV, R. ; GASZNER, M. ; HAIR, A. ; FELSENFELD, G. ; WEST, A.G.** VEZF1 elements mediate protection from DNA methylation. *PLoS Genet*, 2010, vol. 6, e1000804 [0222]
- **LI, C.L. ; XIONG, D. ; STAMATOYANNOPOULOS, G. ; EMERY, D.W.** Genomic and functional assays demonstrate reduced gammaretroviral vector genotoxicity associated with use of the cHS4 chromatin insulator. *Mol Ther*, 2009, vol. 17, 716-724 [0222]
- **LISOWSKI, L. ; SADELAIN, M.** Locus control region elements HS1 and HS4 enhance the therapeutic efficacy of globin gene transfer in beta-thalassemic mice. *Blood*, 2007, vol. 110, 4175-4178 [0222]
- **NAGEL, R.L. ; BOOKCHIN, R.M. ; JOHNSON, J. ; LABIE, D. ; WAJCMAN, H. ; ISAAC-SODEYE, W.A. ; HONIG, G.R. ; SCHILIRO, G. ; CROOKSTON, J.H. ; MATSUTOMO, K.** Structural bases of the inhibitory effects of hemoglobin F and hemoglobin A2 on the polymerization of hemoglobin S. *Proc Natl Acad Sci U S A*, 1979, vol. 76, 670-672 [0222]
- **SADELAIN et al.** *Proc. Nat'l Acad. Sci. (USA)*, 1995, vol. 92, 6728-6732 [0222]
- **ARMSTRONG, J.A. ; EMERSON, B.M.** NFE2 disrupts chromatin structure at human fl-globin locus control region hypersensitive site 2 in vitro. *Mol. Cell. Biol.*, 1996, vol. 16, 5634-5644 [0222]
- **CATERINA, J.J. ; CIAVATTA, D.J. ; DONZE, D. ; BEHRINGER, R.R. ; TOWNES, T.M.** Multiple elements in human fl-globin locus control region 5' HS2 are involved in enhancer activity and position-independent transgene expression. *Nucleic Acids Res.*, 1994, vol. 22, 1006-1011 [0222]
- **MOI, P. ; KAN, Y.W.** Synergistic enhancement of globin gene expression by activator protein-1-like proteins. *Proc. Natl. Acad. Sci. USA*, 1990, vol. 87, 9000-9004 [0222]
- **NEY, P. ; SORRENTINO, B. ; MCDONAGH, K. ; NIENHUIS, A.** Tandem AP-1-binding sites within the human β -globin dominant control region function as an inducible enhancer in erythroid cells. *Genes Dev.*, 1990, vol. 4, 993-1006 [0222]
- **SHIVDASANI, R.A. ; ROSENBLATT, M.F. ; ZUCKER-FRANKLIN, D. ; JACKSON, C.W. ; HUNT, P. ; SARIS, C.J.M. ; ORKIN, S.H.** Transcription factor NF-E2 is required for platelet formation independent of the actions of thrombopoietin/MGDF in megakaryocyte development. *Cell*, 1995, vol. 81, 695-704 [0222]
- **TALBOT, D. ; GROSVELD, F.** The 5'HS2 of the globin locus control region enhances transcription through the interaction of a multimeric complex binding at two functionally distinct NF-E2 binding sites. *EMBO J.*, 1991, vol. 10, 1391-1398 [0222]
- **HARDISON et al.** *Gene*, 1997, vol. 205, 73-94 [0222]
- **ELNITSKI et al.** *The Journal of Biological Chemistry*, 1997, vol. 272 (1), 369-378 [0222]
- **HORAK et al.** *PNAS*, 2002, vol. 99 (5), 2924-2929 [0222]
- **SHIMOTSUMA et al.** *Journal of Biological Chemistry*, 2010, vol. 285 (19), 14495-14503 [0222]